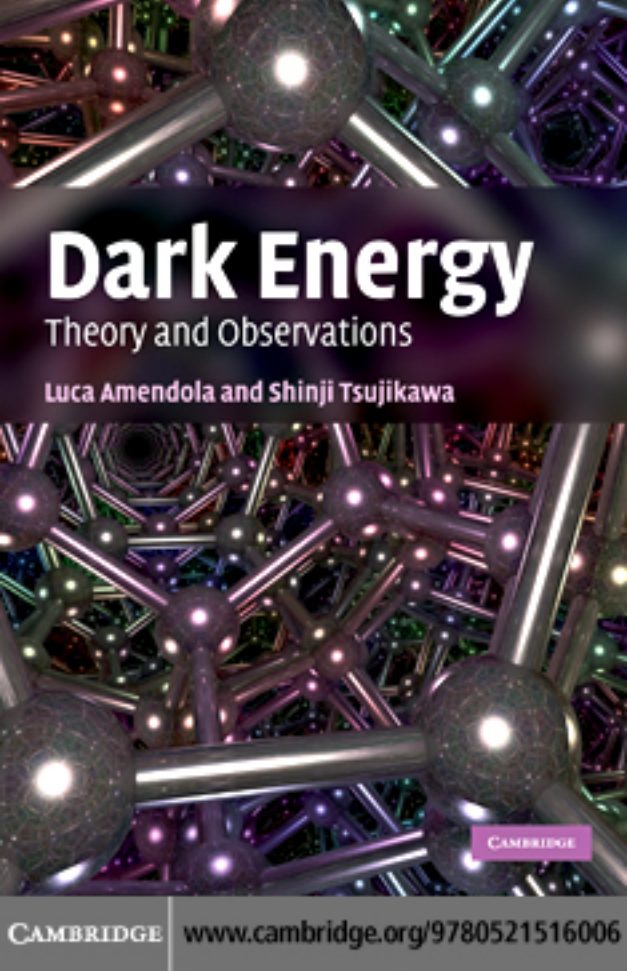


The background of the book cover is a complex, abstract pattern. It features numerous glowing spheres in shades of purple, blue, and white, some of which are larger and more prominent than others. These spheres are interconnected by a dense network of thin, intersecting lines that create a sense of depth and complexity, resembling a molecular structure or a network diagram. The overall color palette is dominated by purples and blues, with bright highlights from the glowing elements.

Dark Energy

Theory and Observations

Luca Amendola and Shinji Tsujikawa

CAMBRIDGE

CAMBRIDGE

www.cambridge.org/9780521516006

This page intentionally left blank

DARK ENERGY

Theory and Observations

Dark energy, the mysterious cause of the accelerating expansion of the Universe, is one of the most important fields of research in astrophysics and cosmology today. Introducing the theoretical ideas, observational methods and results, this textbook is ideally suited to graduate courses on dark energy, and will also supplement advanced cosmology courses.

Providing a thorough introduction to this exciting field, the textbook covers the cosmological constant, quintessence, k-essence, perfect fluid models, extra-dimensional models, and modified gravity. Observational research is reviewed, from the Cosmic Microwave Background to baryon acoustic oscillations, weak lensing, and cluster abundances. Every chapter ends with problems, with full solutions provided, and any calculations are worked through step-by-step.

LUCA AMENDOLA is Professor of Physics at the University of Heidelberg, Germany. He is also an astronomer at the Italian National Institute for Astrophysics, Rome, and has authored more than 100 papers in international journals and conference proceedings.

SHINJI TSUJIKAWA is Associate Professor at Tokyo University of Science in Japan, where he teaches cosmology and relativity. He has published more than 80 papers in international refereed journals and has over 60 collaborators worldwide.

DARK ENERGY

Theory and Observations

LUCA AMENDOLA

*Institut für Theoretische Physik, Universität Heidelberg/
Osservatorio Astronomico di Roma*

SHINJI TSUJIKAWA

*Department of Physics, Faculty of Science,
Tokyo University of Science*



CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,
São Paulo, Delhi, Dubai, Tokyo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org

Information on this title: www.cambridge.org/9780521516006

© L. Amendola and S. Tsujikawa 2010

This publication is in copyright. Subject to statutory exception and to the provision of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published in print format 2010

ISBN-13 978-0-511-75008-3 eBook (NetLibrary)

ISBN-13 978-0-521-51600-6 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy of urls for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Contents

<i>Preface</i>	<i>page</i> ix
<i>List of frequently used symbols</i>	xiii
1 Overview	1
2 Expansion history of the Universe	7
2.1 Friedmann equations	7
2.2 Hubble's law	12
2.3 Matter species in the Universe	13
2.4 Cosmic distances	18
2.5 The equation of state of dark energy	22
2.6 Problems	26
3 Correlation function and power spectrum	27
3.1 The correlation function	27
3.2 The n -point correlation function	30
3.3 The power spectrum	31
3.4 From the power spectrum to the moments	37
3.5 Problems	39
4 Basics of cosmological perturbation theory	40
4.1 Perturbing General Relativity	40
4.2 The Newtonian gauge	42
4.3 Single-fluid model	46
4.4 Scales larger than the horizon	51
4.5 Scales smaller than the Hubble radius	51
4.6 Two-fluid solutions	53
4.7 Velocity field	56
4.8 The redshift distortion	58
4.9 Baryons, photons, and neutrinos	62
4.10 The matter power spectrum	70

4.11	Perturbed photon propagation	76
4.12	Problems	83
5	Observational evidence of dark energy	84
5.1	The age of the Universe	84
5.2	Supernova observations	87
5.3	Cosmic Microwave Background	93
5.4	Baryon acoustic oscillations	102
5.5	Large-scale structure	106
5.6	Problems	108
6	Cosmological constant	109
6.1	Einstein equations with the cosmological constant	110
6.2	History of the cosmological constant	111
6.3	The fine tuning problem	113
6.4	The coincidence problem	114
6.5	Supersymmetric models	116
6.6	Cosmological constant and the anthropic principle	124
6.7	The decoupling of the cosmological constant from gravity	129
6.8	Problems	133
7	Dark energy as a modified form of matter I: Quintessence	134
7.1	Quintessence	135
7.2	Dynamical system approach	138
7.3	Early dark energy	149
7.4	Quintessence potentials in particle physics	153
7.5	Reconstruction of quintessence from observations	163
7.6	Problems	171
8	Dark energy as a modified form of matter II	172
8.1	k-essence	172
8.2	Phantoms	186
8.3	Coupled dark energy	189
8.4	Chameleon scalar fields	205
8.5	Dark energy models with scaling solutions	215
8.6	Unified models of dark energy and dark matter	225
8.7	Future singularities	230
8.8	Problems	233
9	Dark energy as a modification of gravity	234
9.1	$f(R)$ gravity	234
9.2	Scalar-tensor theories	257
9.3	Gauss–Bonnet dark energy models	269
9.4	Braneworld models of dark energy	277
9.5	Problems	283

10	Cosmic acceleration without dark energy	285
10.1	Void models	286
10.2	Backreaction	292
10.3	Problems	295
11	Dark energy and linear cosmological perturbations	296
11.1	Perturbations in a general dark energy cosmology	296
11.2	Perturbations of a scalar field	306
11.3	From dark energy to dark force	309
11.4	A massive dark energy field	313
11.5	Sound speed of a scalar field	314
11.6	Perturbations in modified gravity models	317
11.7	Problems	335
12	Non-linear cosmological perturbations	336
12.1	Second-order perturbations	336
12.2	The bispectrum and the higher-order correction	
	to the power spectrum	341
12.3	Spherical collapse	347
12.4	The mass function of collapsed objects	351
12.5	Dark energy N -body simulations	353
12.6	Problems	355
13	Statistical methods in cosmology	356
13.1	The likelihood function	356
13.2	Model selection	363
13.3	Fisher matrix	367
13.4	The Fisher matrix for the power spectrum	376
13.5	Principal component analysis	379
13.6	Problems	381
14	Future observational constraints on the nature of dark energy	383
14.1	Dark energy and the CMB	383
14.2	Large-scale structure	391
14.3	Growth function	398
14.4	Cosmic shear	403
14.5	Cluster abundances and baryon fraction	409
14.6	Other probes	415
14.7	Problems	426
15	Conclusion and outlook	427
16	Answers to the problems	430
17	Mathematical Appendix	455
	<i>References</i>	457
	<i>Index</i>	485

Preface

Perhaps the first recognition that the matter composing the universe may be different from the one we touch and experience every day has been put in writing by early Greek philosophers and by Aristotle in particular. In his work *On the Heavens*, Aristotle argues that the nature and movement of the stars and planets is so fundamentally different from Earth-like elements that a new substance is required, a “*bodily substance other than the formations we know, prior to them all and more divine than they.*”¹ Later on this cosmic element came to be called *quintessentia*, or fifth element, and drawing on Plato’s classification of the elements a dodecahedron’s figure was associated with it.

More than two thousand years later, astrophysicists have begun to pile up evidence that a new form of matter pervades our Universe. This idea is based on observations that reminds one of Aristotle’s thoughts: the global movement we observe in distant reaches of our cosmos is unexplainable by ordinary matter. All the matter we see on Earth, in the solar system, inside our Galaxy or in similar structures across the Universe has a small or negligible positive pressure and clumps under the influence of gravity. An expanding Universe filled with this form of matter would by necessity slow down. But in 1998, astronomers studying the global expansion by the use of supernovae found that their observed luminosities can be explained only by an accelerated expansion of the Universe. After a full decade of more observations, more analyses and more interpretations, we still cannot find a better explanation than invoking something new, a new force or a new matter, that acts “*on the heavens.*” This new form of matter, called *dark energy*, is what this book is about.

We do not know the nature of dark energy yet. We are beginning to characterize its properties in several ways, from its abundance to its dynamics, but we know still very little about it. The simplest explanation, an energy associated with the

¹ Aristotle, *On the Heavens*, I, 2.

vacuum, was already proposed on a totally different basis by Einstein under the name of universal (Einstein's original term) or cosmological constant² Λ . But the freedom still allowed by observations has unleashed theorists' imagination and many interpretations of dark energy have been advanced. Dark energy is indeed a general label for what we do not know about the large-scale properties of our Universe, its history and its geography.

So why a book on dark energy? One of the immediate consequences of the discovery of the cosmic acceleration and the hypothesis of dark energy has been that astrophysicists and particle physicists, both theorists and experimentalists, have been drawn together into this new field of research, with their own languages and methods. We believe that this has created the need for a resource that allows scholars and students to apprehend the basis of dark energy research in an interdisciplinary way. This book introduces the main theoretical ideas on dark energy and at the same time the basics of the observational methods and results. There are several reviews that cover parts of the dark energy research but not a book that could be used as a starting point to advanced and more topical material.

This book can be used as a companion text for an advanced cosmology course, covering several areas that complement modern cosmology textbooks or as a stand-alone text for graduate or post-graduate courses on dark energy. It is also addressed to newcomers in the field that wish to identify the main lines of the current research. Finally, we have in mind also researchers in the dark energy field who need to explore other sides of the discipline and would like to have a handy reference for many results and topics scattered in the literature. For most of the book we assume knowledge of General Relativity and basic cosmology at the graduate level and little more. Some more advanced sections (especially Sections 6.5 and 7.4) require also a background in quantum field theory but they can be left aside without prejudice. Whenever possible we give a fairly comprehensive review of the tools required for further material, for instance we introduce the concepts of statistics and of cosmological perturbation theory that are needed for understanding the subsequent chapters. We provide 44 fully solved problems with some detailed calculations, which will help the reader to test his/her understanding.

The immense impact on cosmology of the dark energy concept is witnessed by the many projects around the world aiming at collecting more and more data, from large-scale galaxy surveys to weak lensing surveys, from cosmic microwave observations to gamma-ray bursts. The very nature of the issue at stake, the study of a component that determines to a large extent the present and future cosmic dynamics, has generated a great diversity of theoretical and observational approaches. One

² As a curious coincidence, Aristotle first proposed his eternal and incorruptible "cosmic substance" in the book Λ (i.e. the twelfth book) of the *Metaphysics*. "The Lambda Book," as it was called in the middle ages, was singled out as the highest point of Aristotle's metaphysics.

can attack the problem from the point of view of an exotic matter component, or of a non-Einsteinian gravity force, or invoking multidimensional effects. Similarly, one can employ a very diverse array of observations, from standard candles to standard clocks, from supernovae to quasar multiple images. Many of the hypotheses and methods will have a strong impact on cosmology even beyond the dark energy problem. This book tries to extract from this variety the core teachings: methods, suggestions, hypotheses, and techniques that are shaping our knowledge of the cosmos. Many of these, we reckon, will remain with us for many years.

Although the diversity of approaches is one of the hallmarks of dark energy research, we could not possibly cover all the ideas discussed so far. Up to 2009 the number of papers that include the words “dark energy” or “cosmological constant” in the title has been over 3700. We tried to discuss all the driving ideas but not all possible implementations. In doing so we certainly missed some interesting contributions; we apologize in advance to our colleagues.

We thank all our collaborators on the topics in this book, namely, Carlo Baccigalupi, Amedeo Balbi, Marco Baldi, Kazuharu Bamba, Riccardo Barbieri, Bruce A. Bassett, Silvio Bonometto, Stefano Borgani, Robert Brandenberger, Carlo Burigana, Paolo Cabella, Gianluca Calcagni, Gabriela C. Campos, Salvatore Capozziello, Daniela Carturan, Christos Charmousis, Edmund J. Copeland, Pier Stefano Corasaniti, Stephen C. Davis, Antonio De Felice, Cinzia Di Porto, Stephane Fay, Fabio Finelli, Radouane Gannouji, Mohammad Reza Garousi, Maurizio Gasperini, Chao-Qiang Geng, Emanuele Giallongo, Fabio Giori, Burin Gumjudpai, Zong-Kuan Guo, Soo A. Kim, Martin Kunz, Maxim Libanov, Andrew R. Liddle, Roy Maartens, Andrea Macciò, Kei-ichi Maeda, Roberto Mainini, Elisabetta Majerotto, Martin Makler, Matteo Martinelli, Alessandro Melchiorri, Shuntaro Mizuno, Bruno Moraes, David F. Mota, Tapan Naskar, Savvas Nesseris, Shin’ichi Nojiri, Sergei Odintsov, Junko Ohashi, Nobuyoshi Ohta, Sudhakar Panda, Eleftherios Papantonopoulos, David Parkinson, Alessandro Pasqui, Valeria Pettorino, Yun-Song Piao, Federico Piazza, David Polarski, Miguel Quartín, Claudia Quercellini, R. R. R. Reis, Rogerio Rosenfeld, Valery Rubakov, M. Sami, Domenico Sapone, Parampreet Singh, Alexei Starobinsky, Takashi Tamaki, Takayuki Tatekawa, Reza Tavakol, Domenico Tocchini-Valentini, Alexey Toporensky, Peter V. Tretjakov, Roberto Trotta, Kotub Uddin, Carlo Ungarelli, Ioav Waga, David Wands, John Ward, Christof Wetterich, Jun’ichi Yokoyama, and Xinmin Zhang.

Apart from collaborators, we are grateful to many other people with whom we have discussed a lot about dark energy. We also thank Graham Hart and the editorial staff at Cambridge University Press for giving us the opportunity to write this book. Finally we are grateful to our families: Emanuela, Davide, Yasuko, and Masato for their support and love.

L.A. and S.T.

I wish to honor the memory of Franco Occhionero, long-time friend and teacher: you opened to me the magic lamp of cosmology. I also thank the staff and colleagues at the Osservatorio Astronomico di Roma and the Director Emanuele Giallongo, for support, patience, collaboration, help, friendship, and everything that makes work possible, pleasant and productive. Thanks to the University of Heidelberg for the new exciting opportunities they are offering to me. Special thanks to Silvio Bonometto, Roberto Buonanno, Roberto Scaramella and Christof Wetterich for collaboration, advice, support, and friendship. Special thanks to Claudia Quercellini, Ioav Waga, Miguel Quartin, Fabio Finelli, Cinzia di Porto for comments on the draft and for the long-term friendship and collaboration. Thanks also to Adam Amara, Enzo Branchini, Francisco Castander, Gigi Guzzo, Tom Kitching, Rocky Kolb, Martin Kunz, Roberto Maoli, Anais Rassat, Alexandre Refregier, Yun Wang, Jochen Weller, and all the other DUNE/Euclid and NASA/SWG collaborators, too many to be listed here, from whom I learned a lot in terms both of science and of scientific collaboration.

L.A.

I am grateful to Kei-ichi Maeda for giving me a chance to work as a cosmologist, in spite of the fact that I was majoring mathematics before entering a master course of physics in 1996. This was a correct decision because I witnessed enormous progress of observational and theoretical cosmology from the late 1990s, especially in the field of dark energy. I am also thankful to the members of Tokyo University of Science and Gunma National College of Technology, especially to Antonio De Felice, Hitoshi Fujiwara and Junko Ohashi, for their kind support and help. I also thank Bruce A. Bassett and Roy Maartens, who kindly helped me in many aspects during my stay in Portsmouth in 2003. I am grateful to M. Sami for a long-term collaboration about dark energy.

S.T.

Frequently used symbols

Symbol	Definition
G	Gravitational constant ($G = 6.67 \times 10^{-8} \text{ cm}^3 \text{ g}^{-1} \text{ sec}^{-2}$)
m_{pl}	Planck mass ($m_{\text{pl}} = 1/\sqrt{G} = 1.2211 \times 10^{19} \text{ GeV}$)
M_{pl}	Reduced Planck mass ($M_{\text{pl}} = 1/\sqrt{8\pi G} = 2.4357 \times 10^{18} \text{ GeV}$)
κ	$\sqrt{8\pi G}$
a	Scale factor of the Universe (with the present value $a_0 = 1$)
t	Cosmic time
η	Conformal time: $\eta = \int a^{-1} dt$
N	Number of e-foldings: $N = \ln a$
$\cdot$	Derivative with respect to t
$'$	Derivative with respect to η (or $N = \ln a$ in Chapters 11 and 12)
z	Redshift: $z = a_0/a - 1$
d_A, d_L	Angular diameter distance, luminosity distance
$H, \mathcal{H}$	Hubble parameter: $H = \dot{a}/a$, conformal Hubble parameter $\mathcal{H} = aH$
H_0, h	Present Hubble parameter: $H_0 = 100 h \text{ km sec}^{-1} \text{ Mpc}^{-1}$
$E(z)$	Hubble parameter normalized by H_0 : $E(z) = H(z)/H_0$
ρ	(Energy) Density
P	Pressure
w	Equation of state: $w = P/\rho$
w_{eff}	Effective or total equation of state: $w_{\text{eff}} = -1 - 2\dot{H}/(3H^2)$
K	Curvature of the Universe
R	Ricci scalar
$\Omega^{(0)}$	Density parameter at the present epoch ($z = 0$)
c_s	Sound speed
r_s	Sound horizon: $r_s(\eta) = \int_0^\eta d\tilde{\eta} c_s(\tilde{\eta})$
Φ, Ψ	Gravitational potentials
T	Temperature

Symbol	Definition
Θ	Temperature perturbations: $\Theta = \delta T/T$
k	Comoving wavenumber
$P(k)$	Power spectrum of perturbations
ℓ	Spherical harmonic multipoles
C_ℓ	Multipole power spectrum
$\mathcal{R}$	CMB shift parameter
δ	Density contrast
D	Growth function
b	Bias (ratio of galaxy to total matter perturbations)
Λ	Cosmological constant
S	Action
$g_{\mu\nu}$	Metric
$G_{\mu\nu}$	Einstein tensor
$T_{\mu\nu}$	Energy-momentum tensor
ϕ	Scalar field
$V(\phi)$	Scalar-field potential in the Einstein frame
$U(\phi)$	Scalar-field potential in the Jordan frame
$\mathcal{L}$	Lagrangian density (also log-likelihood)
λ	Slope of the potential defined by $\lambda = -V_{,\phi}/(\kappa V)$
$\hat{\lambda}$	Dimensionless perturbation scale $\hat{\lambda} = \mathcal{H}/k$
X	Kinetic energy: $X = -(1/2)g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi$
Q	Coupling between a scalar field ϕ and non-relativistic matter
ω_{BD}	Brans–Dicke parameter
R_{GB}^2	Gauss–Bonnet term

1

Overview

From the observational data of Supernovae Type Ia (SN Ia) accumulated by the year 1998, Riess *et al.* [1] in the High-redshift Supernova Search Team and Perlmutter *et al.* [2] in the Supernova Cosmology Project Team independently reported that the present Universe is accelerating. The source for this late-time cosmic acceleration was dubbed “dark energy.” Despite many years of research (see e.g., the reviews [3, 4, 5, 6, 7]) its origin has not been identified yet. Dark energy is distinguished from ordinary matter species such as baryons and radiation, in the sense that it has a negative pressure. This negative pressure leads to the accelerated expansion of the Universe by counteracting the gravitational force. The SN Ia observations have shown that about 70% of the present energy of the Universe consists of dark energy.

The expression “dark energy” may be somewhat confusing in the sense that a similar expression, “dark matter,” has been used to describe a pressureless matter (a non-relativistic matter) that interacts very weakly with standard matter particles. The existence of dark matter was already pointed out by Zwicky in the 1930s by comparing the dispersion velocities of galaxies in the Coma cluster with the observable star mass. Since dark matter does not mediate the electromagnetic force, its presence is mainly inferred from gravitational effects on visible matter. Dark matter can cluster by gravitational instability (unlike standard dark energy) so that local structures have been formed in the Universe. In fact it is observationally known that dark matter has played a crucial role for the growth of large-scale structure such as galaxies and clusters of galaxies. The energy fraction of dark matter in the present universe is about 25%, whereas that of baryons is about 4%. The black body radiation, which dominated over the other matter components in the past, shares only about 0.005% of the present total energy density.

In modern cosmology it is believed that another cosmic acceleration called “inflation” occurred in the very early Universe prior to the radiation-dominated

epoch. The idea of inflation was originally proposed in the early 1980s by a number of people [8, 9, 10, 11] to solve several cosmological problems such as the flatness and horizon problems. Inflation also provides a causal mechanism for the origin of large-scale structure in the Universe. The temperature anisotropies of Cosmic Microwave Background (CMB) observed by the Cosmic Background Explorer (COBE) in 1992 [12] showed that the fluctuation spectrum is nearly scale-invariant.¹ This is consistent with theoretical predictions of the power spectrum of density perturbations originated from quantum fluctuations of a scalar field generated during inflation. After 2003, the Wilkinson Microwave Anisotropy Probe (WMAP) group has provided high-precision observational data of CMB anisotropies [13, 14, 15]. This has given strong support for the existence of an inflationary period as well as dark energy.

After the end of inflation the Universe entered the radiation-dominated epoch during which light elements such as helium and deuterium were formed. Since the energy density of radiation decreases faster than that of non-relativistic matter such as dark matter and baryons, the radiation-dominated era is eventually followed by the matter-dominated epoch around the redshift $z = 3000$. The temperature anisotropies observed by COBE and WMAP occur on the last scattering surface at which electrons were trapped by hydrogen to form atoms. After this *decoupling epoch* photons can freely move to us without experiencing Thomson scattering. The decoupling corresponds to the redshift $z \simeq 1090$. According to the WMAP 5-year data [15], the energy components at the decoupling epoch are dark matter (63%), radiations (25%) [photons (15%) and neutrinos (10%)], and baryons (12%) with at most a tiny amount of dark energy. We will often make reference to the cosmological parameters measured by WMAP in the course of this book.

The formation of structure (galaxies, clusters) started in the matter-dominated epoch, i.e. when the pressureless dark matter began to dominate the total energy density of the Universe. Baryons also contribute to the formation of large-scale structure to some extent. During the matter era the energy density of dark energy needs to be suppressed compared to that of dark matter in order to allow sufficient growth of large-scale structure. If dark energy couples to dark matter with some interaction (as in the coupled quintessence scenario [16, 17]), then dark energy also affects the past expansion history of the Universe as well as the structure formation. It is possible to place bounds on the strength of such couplings from the observations of CMB and of galaxy clustering. In addition to the experiments of direct and indirect dark matter search (see e.g., [18, 19, 20, 21]), like those at the Large Hadron Collider (LHC) at CERN and in underground, ground, and space

¹ J. Mather and G. Smoot won the Nobel Prize in 2006 for the measurement of the black body spectrum and the discovery of the temperature anisotropy of CMB.

facilities, cosmological observations will shed light on the relation between dark matter and dark energy.

While the energy density of dark matter evolves as $\rho_m \propto a^{-3}$ (a is the scale factor of an expanding Universe), the dark energy density is nearly constant in time ($\rho_{DE} \propto a^{-n}$ with n probably close to 0). Hence the latter energy density eventually catches up with the former. The onset of the cosmic acceleration occurs around the redshift $z \sim 1$, although there is still uncertainty for its precise value because of the model-dependence. We live then in a special epoch of the cosmic acceleration in the long expansion history of the Universe. The problem why the accelerated expansion of the Universe started around today is often called the “coincidence problem.” The standard radiation- and matter-dominated eras are sandwiched by two periods of cosmic acceleration – inflation and dark energy.

The simplest candidate for dark energy is the so-called cosmological constant Λ , whose energy density remains constant [22]. Originally the cosmological constant was introduced by Einstein in 1917 to realize a static Universe in the framework of General Relativity [23]. In fact the Einstein equations allow the freedom to add the constant Λ term. The cosmological constant works as a negative pressure against gravity so that the two effects can balance each other. However, after the discovery of the expansion of the Universe by Hubble from the measurement of recession speeds of distant galaxies, Einstein abandoned the idea of adding the Λ term to the Einstein equations. At the late stage of his career, he regretted having introduced Λ as his “*biggest blunder*” (or so is told by George Gamow). In fact, there was nothing to regret: after 1998 the cosmological constant revived again as a form of dark energy responsible for the late-time acceleration of the Universe.

From the viewpoint of particle physics, the cosmological constant appears as vacuum energy density. If we sum up zero-point energies of all normal modes of some field and take the cut-off scale of the momentum at the Planck scale, the vacuum energy density is estimated to be $\rho_{vac} \simeq 10^{74} \text{ GeV}^4$. This is much larger than the observed value of dark energy: $\rho_\Lambda \simeq 10^{-47} \text{ GeV}^4$. If vacuum energy with an energy density of the order of $\rho_{vac} \simeq 10^{74} \text{ GeV}^4$ was present in the past, the Universe would have entered an eternal stage of cosmic acceleration already in the very early Universe. This is of course problematic because the success of the big bang cosmology based on the presence of radiation and matter epochs is completely destroyed. Hence the problem of the large vacuum energy density was known long before the discovery of dark energy in 1998.

If the cosmological constant is responsible for the present cosmic acceleration, we need to find a mechanism to obtain the tiny value of Λ consistent with observations. A lot of efforts have been made in this direction under the framework of particle physics. For example, the recent development of string theory shows that it is possible to construct de Sitter vacua by compactifying extra dimensions in the

presence of fluxes with an account of non-perturbative corrections [24]. The fact that there is a huge number of different choices of fluxes gives rise to the so-called “string landscape” with more than 10^{500} vacua [25]. Some scientists argued that only the vacuum whose energy density is of the order of the present cosmological density can sustain life or complexity and this explains why we live in a low- Λ world. This anthropic argument is, to say the least, highly controversial.

If the origin of dark energy is not the cosmological constant, one may seek for some alternative models to explain the cosmic acceleration today. Basically there are two approaches to construct models of dark energy other than the cosmological constant.

The first approach is to modify the right-hand side (r.h.s.) of the Einstein equations given in Eq. (2.8) by considering specific forms of the energy-momentum tensor $T_{\mu\nu}$ with a negative pressure. The representative models that belong to this class are the so-called cosmon or quintessence [26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41], k-essence [42, 43, 44], and perfect fluid models [45, 46]. The quintessence makes use of scalar fields with slowly varying potentials, whereas in k-essence it is the scalar-field kinetic energy that drives the acceleration. The perfect fluid models are based on a perfect fluid with a specific equation of state such as the Chaplygin gas model [45] and its generalizations [46]. There have been many attempts to construct scalar-field models of dark energy based on particle physics (see Refs. [47, 48, 49, 50, 51, 52] for early works). In the context of inflation, since the associated energy scale is high, it is natural for scalar fields to be responsible for the acceleration of the Universe. The situation is different for dark energy – its energy scale is too low compared to typical scales appearing in particle physics. Moreover, the field potentials need to be sufficiently flat so that the field evolves slowly enough to drive the present cosmic acceleration. This demands that the field mass is extremely light ($m_\phi \simeq 10^{-33}$ eV) relative to typical mass scales appearing in particle physics. It would be expected that this light scalar field should mediate long-range forces with ordinary matter [36]. Such couplings need to be suppressed in order to be consistent with a number of local gravity experiments. In spite of the above-mentioned difficulties it is not hopeless to construct viable scalar-field dark energy models in the framework of particle physics.

The second approach for the construction of dark energy models is to modify the left-hand side (l.h.s.) of the Einstein equations (2.8). The representative models that belong to this class (that we denote “modified gravity”) are the so-called $f(R)$ gravity [53, 54, 55], scalar-tensor theories [56, 57, 58, 59, 60], and braneworld models [61, 62]. The cosmological constant scenario (in other words, the “ Λ -Cold-Dark-Matter (Λ CDM) model”) corresponds to the Lagrangian density $f(R) = R - 2\Lambda$, where R is the Ricci scalar. A possible modification of the Λ CDM is described by a non-linear Lagrangian density f in terms of R , which is called

$f(R)$ gravity. Scalar-tensor theories correspond to theories in which the Ricci scalar R couples to a scalar field ϕ with a coupling of the form $F(\phi)R$. They include Brans–Dicke theory [63] and dilaton gravity [64] as specific cases. In the braneworld models proposed by Dvali, Gabadadze, and Porrati (DGP) [61] the late-time acceleration of the Universe can be realized as a result of the gravitational leakage from a 3-dimensional surface (3-brane) to a 5-th extra dimension on Hubble distances. Generally we require that modified gravity models satisfy local gravity constraints as well as conditions for the cosmic acceleration preceded by the matter-dominated epoch. In this sense modified gravity models are typically more strongly constrained than modified matter models from gravitational experiments and cosmological observations.

It is important to realize however that the two approaches, which we denote as modified matter and modified gravity, are not fundamentally different, at least if for a moment we do not consider their quantum field implications. From the viewpoint of classical General Relativity (which is all that matters for most of cosmology), one can always rephrase one into the other by defining a suitable conserved energy-momentum tensor that equals the Einstein tensor.

In order to distinguish this variety of models of dark energy, it is important to place constraints by using observational data such as SN Ia, CMB, and large-scale structure (LSS). Usually the equation of state of dark energy, $w_{\text{DE}} \equiv P_{\text{DE}}/\rho_{\text{DE}}$, where P_{DE} is the pressure and ρ_{DE} is the energy density, is a good measure to describe the property of dark energy at background level. In the case of the cosmological constant we have $P_{\text{DE}} = -\rho_{\text{DE}}$ and hence $w_{\text{DE}} = -1$. In other models of dark energy the equation of state w_{DE} generally varies in time. Perhaps the first task of dark energy research is to detect deviations from the value $w_{\text{DE}} = -1$ in order to find whether dark energy can be identified with the cosmological constant or not.

The SN Ia observations have provided information of the cosmic expansion history around the redshift $z \lesssim 2$ by the measurement of luminosity distances of the sources. The presence of dark energy leads to a shift of the position of acoustic peaks in CMB anisotropies as well as a modification of the large-scale CMB spectrum through the so-called integrated Sachs–Wolfe effect. Although the CMB data alone are not sufficient to place strong constraints on dark energy, the combined analysis of SN Ia and CMB can provide tight bounds on the equation of state w_{DE} and the present energy fraction $\Omega_{\text{DE}}^{(0)}$ of dark energy [15]. The distribution of large-scale clustering of galaxies in the sky also provides additional information on the properties of dark energy [65, 66, 67]. In 2005 the detection of a peak of baryon acoustic oscillations (BAO) was reported by Eisenstein *et al.* [68] at the average redshift $z = 0.35$ from the observations of luminous red galaxies in the Sloan Digital Sky Survey. This has also given us another independent test of dark

energy. From the combined analysis of SN Ia, CMB, and BAO, the WMAP group [15] obtained the bound $-1.097 < w_{\text{DE}} < -0.858$ at the 95% confidence level assuming a constant equation of state. The cosmological constant ($w_{\text{DE}} = -1$) is well consistent with the current observational data while some dark energy models have been already excluded from observations.

In future observations it is expected that other observational data such as weak gravitational lensing and gamma ray bursts will shed light on the nature of dark energy. Confirming Λ CDM or detecting deviations from it would be an extremely important step towards understanding the origin of dark energy.

Units and conventions

Throughout this book we use units such that $c = \hbar = k_{\text{B}} = 1$, where c is the speed of light, $\hbar$ is reduced Planck's constant, and k_{B} is Boltzmann's constant. We reinsert these symbols when the discussion needs it. In these units everything can be expressed in terms of a single unit, e.g., time, length, or energy. The gravitational constant G is related to the Planck mass $m_{\text{pl}} = 1.2211 \times 10^{19} \text{ GeV}$ via $G = 1/m_{\text{pl}}^2$ and the reduced Planck mass $M_{\text{pl}} = 2.4357 \times 10^{18} \text{ GeV}$ via $\kappa^2 \equiv 8\pi G = 1/M_{\text{pl}}^2$, respectively. We adopt the metric signature $(-, +, +, +)$. We list frequently used symbols after the Table of Contents.

2

Expansion history of the Universe

Standard hot big bang cosmology is based on the cosmological principle, which states that the Universe is homogeneous and isotropic at least on large scales. This is supported by a number of observations, such as the CMB photons coming from different parts of the sky with almost the same temperature. The past cosmic expansion history is recovered by solving the Einstein equations in the background of the homogeneous and isotropic Universe. Of course we observe inhomogeneities and irregularities in the local region of the Universe such as stars and galaxies. These inhomogeneities have grown in time through gravitational instability from a matter distribution that was more homogeneous in the past. Then the inhomogeneities can be regarded as small perturbations evolving on the background (homogeneous) Universe.

In this chapter we provide the basic tools to understand the expansion history of the Universe. We also introduce a number of cosmic distances often used to put observational constraints on dark energy.

2.1 Friedmann equations

The line-element that describes a 4-dimensional homogeneous and isotropic spacetime is called Friedmann–Lemaître–Robertson–Walker (FLRW) spacetime and is given by

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = -dt^2 + a^2(t) d\sigma^2, \quad (2.1)$$

where $g_{\mu\nu}$ is a metric tensor, $a(t)$ is a scale factor with cosmic time t , and $d\sigma^2$ is the time-independent metric of the 3-dimensional space with a constant curvature K :

$$d\sigma^2 = \gamma_{ij} dx^i dx^j = \frac{dr^2}{1 - Kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2). \quad (2.2)$$

Here $K = +1, -1, 0$ correspond to closed, open, and flat geometries, respectively. We have used polar coordinates $(x^1, x^2, x^3) = (r, \theta, \phi)$ with $\gamma_{11} = (1 - Kr^2)^{-1}$, $\gamma_{22} = r^2$, and $\gamma_{33} = r^2 \sin^2 \theta$. In Eq. (2.1) the Greek indices μ and ν run from 0 to 3, whereas in Eq. (2.2) the Latin indices i and j run from 1 to 3; the same convention applies to the whole book except when indicated otherwise. We follow Einstein's convention that the terms with same upper and lower indices are summed over. See the book of Weinberg [69] for the derivation of the metric (2.1) from a maximally symmetric spacetime. In addition to the cosmic time t , we also introduce the conformal time η defined by

$$\eta \equiv \int a^{-1} dt. \quad (2.3)$$

The dynamical equations of motion in the expanding Universe can be derived from the Einstein equations by the following steps. From the metric $g_{\mu\nu}$ we obtain the Christoffel symbol:

$$\Gamma_{\nu\lambda}^{\mu} = \frac{1}{2} g^{\mu\alpha} (g_{\alpha\nu,\lambda} + g_{\alpha\lambda,\nu} - g_{\nu\lambda,\alpha}), \quad (2.4)$$

where $g_{\alpha\nu,\lambda} \equiv \partial g_{\alpha\nu} / \partial x^{\lambda}$. Note that $g_{\alpha\nu}$ satisfies the relation $g^{\mu\alpha} g_{\alpha\nu} = \delta_{\nu}^{\mu}$, where δ_{ν}^{μ} is Kronecker's delta ($\delta_{\nu}^{\mu} = 1$ for $\mu = \nu$ and $\delta_{\nu}^{\mu} = 0$ for $\mu \neq \nu$). The Ricci tensor is defined by

$$R_{\mu\nu} = \Gamma_{\mu\nu,\alpha}^{\alpha} - \Gamma_{\mu\alpha,\nu}^{\alpha} + \Gamma_{\mu\nu}^{\alpha} \Gamma_{\alpha\beta}^{\beta} - \Gamma_{\mu\beta}^{\alpha} \Gamma_{\alpha\nu}^{\beta}. \quad (2.5)$$

The contraction of the Ricci tensor gives the Ricci scalar (scalar curvature)

$$R = g^{\mu\nu} R_{\mu\nu}. \quad (2.6)$$

We can then evaluate the Einstein tensor

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R. \quad (2.7)$$

The cosmological dynamics can be obtained by solving the Einstein equations

$$G_{\nu}^{\mu} = 8\pi G T_{\nu}^{\mu}, \quad (2.8)$$

where T_{ν}^{μ} is the energy-momentum tensor of matter components. The l.h.s. of Eq. (2.8) characterizes the geometry of spacetime, whereas the r.h.s. describes energies and momenta of matter components. In the cosmological setting the cosmic expansion rate is determined by specifying the properties of matter in the Universe.

For the FLRW metric (2.1) the non-vanishing components of Christoffel symbols are

$$\Gamma_{ij}^0 = a^2 H \gamma_{ij}, \quad \Gamma_{0j}^i = \Gamma_{j0}^i = H \delta_j^i, \quad (2.9)$$

$$\Gamma_{11}^1 = \frac{Kr}{1 - Kr^2}, \quad \Gamma_{22}^1 = -r(1 - Kr^2), \quad \Gamma_{33}^1 = -r(1 - Kr^2) \sin^2 \theta, \quad (2.10)$$

$$\Gamma_{33}^2 = -\sin \theta \cos \theta, \quad \Gamma_{12}^2 = \Gamma_{21}^2 = \Gamma_{13}^3 = \Gamma_{31}^3 = \frac{1}{r}, \quad \Gamma_{23}^3 = \Gamma_{32}^3 = \cot \theta, \quad (2.11)$$

where

$$H \equiv \dot{a}/a. \quad (2.12)$$

A dot represents a derivative with respect to cosmic time t . The quantity H , called the Hubble parameter, describes the expansion rate of the Universe. The Christoffel symbols given in Eqs. (2.10) and (2.11) correspond to those for the three-dimensional metric (2.2) with the curvature K .

From Eqs. (2.5) and (2.6) the Ricci tensor and the scalar curvature are

$$R_{00} = -3(H^2 + \dot{H}), \quad R_{0i} = R_{i0} = 0, \quad R_{ij} = a^2(3H^2 + \dot{H} + 2K/a^2)\gamma_{ij}, \quad (2.13)$$

$$R = 6(2H^2 + \dot{H} + K/a^2). \quad (2.14)$$

From Eq. (2.7) together with the relation $G_\nu^\mu = g^{\mu\alpha} G_{\alpha\nu}$, the Einstein tensor is

$$G_0^0 = -3(H^2 + K/a^2), \quad G_i^0 = G_0^i = 0, \quad G_j^i = -(3H^2 + 2\dot{H} + K/a^2)\delta_j^i. \quad (2.15)$$

In the FLRW spacetime the energy-momentum tensor of the background matter is restricted to take the perfect fluid form:

$$T_\nu^\mu = (\rho + P)u^\mu u_\nu + P\delta_\nu^\mu, \quad (2.16)$$

where $u^\mu = (-1, 0, 0, 0)$ is the four-velocity of the fluid in comoving coordinates, and ρ and P are functions of t . The (00) and (ij) components of T_ν^μ are $T_0^0 = -\rho$ and $T_j^i = P\delta_j^i$. Then ρ and P have the meaning of an energy density and a pressure, respectively. Since we are using the unit $c = 1$, the density ρ is not particularly distinguished from the energy density ρc^2 . From the (00) and (ii) components of the Einstein equations (2.8) we obtain

$$H^2 = \frac{8\pi G}{3}\rho - \frac{K}{a^2}, \quad (2.17)$$

$$3H^2 + 2\dot{H} = -8\pi G P - \frac{K}{a^2}. \quad (2.18)$$

Eliminating the term K/a^2 gives

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P). \quad (2.19)$$

Multiplying Eq. (2.17) by a^2 , differentiating and using Eq. (2.19), we find

$$\dot{\rho} + 3H(\rho + P) = 0. \quad (2.20)$$

The Einstein tensor satisfies the Bianchi identities

$$\nabla_\mu G^\mu_\nu \equiv \frac{\partial G^\mu_\nu}{\partial x^\mu} + \Gamma^\mu_{\alpha\mu} G^\alpha_\nu - \Gamma^\alpha_{\nu\mu} G^\mu_\alpha = 0, \quad (2.21)$$

where ∇_μ denotes the covariant derivative. Sometimes we use also the symbol “ $_{;\mu}$ ” to represent the covariant derivative. From the Einstein equations (2.8) it follows that $\nabla_\mu T^\mu_\nu = 0$, which gives the same equation as (2.20) in the FLRW background (see problem 2.1). Hence Eq. (2.20) is called the conservation or continuity equation.

Equation (2.17) can be written in the form:

$$\Omega_M + \Omega_K = 1, \quad (2.22)$$

where

$$\Omega_M \equiv \frac{8\pi G\rho}{3H^2}, \quad \Omega_K \equiv -\frac{K}{(aH)^2}. \quad (2.23)$$

We often refer to the present values of the density parameters. For relativistic particles, non-relativistic matter, dark energy, and curvature, we have, respectively

$$\Omega_r^{(0)} = \frac{8\pi G\rho_r^{(0)}}{3H_0^2}, \quad \Omega_m^{(0)} = \frac{8\pi G\rho_m^{(0)}}{3H_0^2}, \quad \Omega_{\text{DE}}^{(0)} = \frac{8\pi G\rho_{\text{DE}}^{(0)}}{3H_0^2}, \quad \Omega_K^{(0)} = -\frac{K}{(a_0 H_0)^2}. \quad (2.24)$$

When we wish to identify the electromagnetic radiation, rather than all the relativistic particles, we use the subscript γ . When we need to distinguish between (cold) dark matter and baryons we use the subscripts c and b , respectively.¹ When we refer to the cosmological constant, we also use the subscript $\wedge$ instead of DE. Finally, sometimes we use M to denote a generic matter component.

If the expansion of the Universe is decelerated (i.e. $\ddot{a} < 0$) then the curvature term $|\Omega_K|$ continues to increase (because the term $aH (= \dot{a})$ decreases), apart from the case where the Universe is exactly flat ($K = 0$) from the very beginning. The WMAP 5-year data [15] constrain the curvature of the present Universe to be $-0.0175 < \Omega_K^{(0)} < 0.0085$ at the 95% confidence level. We need a phase of

¹ See Section 2.3 for the definition of cold dark matter.

cosmic acceleration ($\ddot{a} > 0$) to reduce $|\Omega_K|$ in the past cosmic expansion history, unless the initial state of the Universe is extremely close to the flat one. In order to realize the present level of flatness of the Universe, we require, prior to the radiation-dominated epoch, a phase of cosmic inflation during which the scale factor changes by more than e^{70} times [70].

Let us consider the case in which the Universe is dominated by a single component with an equation of state defined by

$$w \equiv P/\rho. \quad (2.25)$$

If w is a constant, one can analytically find the evolution of ρ and a for the flat Universe ($K = 0$). Solving Eqs. (2.17) and (2.20) in this case, we obtain the following solutions

$$\rho \propto a^{-3(1+w)}, \quad a \propto (t - t_i)^{2/(3(1+w))}, \quad (2.26)$$

where t_i is a constant. Since from statistical mechanics we know that radiation has the equation of state $w = 1/3$ (as we will see later), it follows that the cosmic evolution during the radiation-dominated epoch is given by $\rho \propto a^{-4}$ and $a \propto (t - t_i)^{1/2}$. Non-relativistic matter corresponds to the case with a negligible pressure relative to its energy density, i.e. $w \simeq 0$. Then the evolution during the matter-dominated era is given by $\rho \propto a^{-3}$ and $a \propto (t - t_i)^{2/3}$.

In order to give rise to the cosmic acceleration we require $\ddot{a} > 0$ in Eq. (2.19), i.e.

$$P < -\rho/3 \quad \rightarrow \quad w < -1/3, \quad (2.27)$$

where ρ is assumed to be positive. The fact that the negative pressure leads to the cosmic acceleration may look counter-intuitive. In Newtonian gravity the pressure is related to a force associated with a local potential that depends on the position in space. In the homogeneous and isotropic Universe such a local potential is absent, which means that there is no Newtonian-analog pressure. In other words, the time-dependent pressure $P(t)$ in the FLRW spacetime appears only in General Relativity. The mechanisms that generate this negative pressure and the cosmic acceleration are the main topic of this book.

When $w = -1$, i.e. $P = -\rho$, it follows from Eq. (2.20) that ρ is a constant. This case corresponds indeed to the so-called cosmological constant. Since H is constant in the flat Universe ($K = 0$), the scale factor evolves exponentially: $a \propto \exp(Ht)$. The cosmological constant cannot be responsible for inflation in the early Universe because otherwise the accelerated expansion would not end. However, it is possible that the cosmological constant is responsible for dark energy because the current cosmic acceleration might indeed continue without end.

2.2 Hubble's law

In the 1920s Slipher and Hubble found that the observed wavelength λ_0 of absorption lines of distant galaxies is larger than the wavelength λ in the rest frame [71]. This is due to the fact that the wavelength is stretched in proportion to the scale factor in an expanding Universe. In order to quantify this effect, we introduce the redshift

$$z \equiv \frac{\lambda_0}{\lambda} - 1 = \frac{a_0}{a} - 1, \quad (2.28)$$

where the present epoch corresponds to $z = 0$. In the following we take the present scale factor a_0 to be unity unless otherwise stated. As we go back to the past, z gets larger. As long as the recessional velocity v of an object is much smaller than the speed of light c we have $\lambda_0 \simeq (1 + v/c)\lambda$ from the Doppler effect, giving

$$z \simeq v/c. \quad (2.29)$$

In an expanding Universe a physical distance $\mathbf{r}$ from an observer (at the origin) to an object is given by $\mathbf{r} = a(t)\mathbf{x}$, where $\mathbf{x}$ denotes the comoving distance. For objects moving with the Hubble flow, the comoving distance remains constant. Taking the derivative of the equation $\mathbf{r} = a(t)\mathbf{x}$ with respect to t , we obtain

$$\dot{\mathbf{r}} = H\mathbf{r} + a\dot{\mathbf{x}}. \quad (2.30)$$

The velocity $\mathbf{v}_H \equiv H\mathbf{r}$ appears because of the presence of the cosmic expansion. On the other hand, the velocity $\mathbf{v}_p \equiv a\dot{\mathbf{x}}$, called *peculiar velocity*, describes the movement of an object with respect to the local Hubble flow. The speed of the object along the direction from the observer to the object is given by

$$v \equiv \dot{\mathbf{r}} \cdot \mathbf{r}/r = Hr + \mathbf{v}_p \cdot \mathbf{r}/r, \quad (2.31)$$

where $r \equiv |\mathbf{r}|$.

In most cases the peculiar velocity of galaxies does not exceed 10^6 m/s. Under the condition that the term $\mathbf{v}_p \cdot \mathbf{r}/r$ is negligible relative to the term Hr , we obtain

$$v \simeq H_0 r. \quad (2.32)$$

Here we have replaced H for the present value H_0 , which is justified in small redshift regions ($z \ll 1$). In 1929, Hubble reported the law (2.32) by plotting the recessional velocity v versus the distance r . His data were scarce, shallow, and noisy, but Hubble concluded correctly that the universe was expanding.

The Hubble parameter H_0 (Hubble constant) is usually written as

$$H_0 = 100 h \text{ km sec}^{-1} \text{ Mpc}^{-1} = 2.1332h \times 10^{-42} \text{ GeV}, \quad (2.33)$$

where

$$1 \text{ Mpc} = 3.08568 \times 10^{24} \text{ cm} = 3.26156 \times 10^6 \text{ light years}. \quad (2.34)$$

Note that h describes the uncertainty on the value H_0 . The observations of the Hubble Key Project constrain this value to be [72]

$$h = 0.72 \pm 0.08. \quad (2.35)$$

Originally Hubble derived a much larger value, $H_0 \sim 500 \text{ km sec}^{-1} \text{ Mpc}^{-1}$, due to the uncertainty of the measurement of distances at that time. We define the Hubble time

$$t_H \equiv 1/H_0 = 9.78 \times 10^9 h^{-1} \text{ years}, \quad (2.36)$$

which is a rough measure of the age of the Universe. The present Hubble radius is defined by

$$D_H \equiv \frac{c}{H_0} = 2998 h^{-1} \text{ Mpc}, \quad (2.37)$$

which corresponds roughly to the largest scale we can observe now.

It is also convenient to introduce the critical density

$$\rho_c^{(0)} \equiv \frac{3H_0^2}{8\pi G} = 1.88 h^2 \times 10^{-29} \text{ g cm}^{-3}, \quad (2.38)$$

which represents the averaged cosmological density in the Universe today. Note that we have used Eqs. (2.33) and (2.34) together with the value $G = 6.67 \times 10^{-8} \text{ cm}^3 \text{ g}^{-1} \text{ sec}^{-2}$ to obtain the numerical value in Eq. (2.38). The critical density (2.38) is very small compared to densities in the local structure of the Universe ($\rho \simeq 5 \text{ g/cm}^3$ for Earth and $\rho \simeq 10^{-24} \text{ g/cm}^3$ for the homogeneous baryon/dark matter density in our Galaxy). An even smaller fraction is responsible for the present accelerated expansion of the Universe.

2.3 Matter species in the Universe

Let us consider matter species in the Universe. They are broadly classified into relativistic particles, non-relativistic matter, and dark energy. Another component, presumably a scalar field, dominated during the period of inflation in the early Universe. In the following we shall first review the equilibrium thermodynamics of relativistic and non-relativistic particles and then proceed to the brief thermal history of the Universe.

Let us consider a particle with momentum $\mathbf{p}$ and mass m . From Special Relativity the energy of this particle is $E = \sqrt{p^2 + m^2}$, where $p \equiv |\mathbf{p}|$. The phase space occupancy $f(\mathbf{p})$ in equilibrium at temperature T is given by the following distribution function

$$f(\mathbf{p}) = \frac{1}{\exp[(E - \mu)/T] \pm 1}, \quad (2.39)$$

where μ is the chemical potential of each species. The plus and minus signs represent the Fermi–Dirac distribution and the Bose–Einstein distribution, respectively.

Generally the distribution function f depends on the position $\mathbf{x}$ of the species, but Eq. (2.39) only depends on $p \equiv |\mathbf{p}|$ because of the homogeneity of the Universe. Since the minimum volume of phase space in terms of $\mathbf{x}$ and $\mathbf{p}$ is given by $(2\pi\hbar)^3$ due to Heisenberg's principle, the number of phase space elements is $d^3x d^3p/(2\pi\hbar)^3$. Then the energy density ρ and the pressure P with g_* internal degrees of freedom are [73, 74]

$$\rho = g_* \int \frac{d^3p}{(2\pi\hbar)^3} E(p) f(p) = \frac{g_*}{2\pi^2} \int_m^\infty dE \frac{(E^2 - m^2)^{1/2}}{\exp[(E - \mu)/T] \pm 1} E^2, \quad (2.40)$$

$$\begin{aligned} P &= g_* \int \frac{d^3p}{(2\pi\hbar)^3} \frac{pv}{3} f(p) = g_* \int \frac{d^3p}{(2\pi\hbar)^3} \frac{p^2}{3E} f(p) \\ &= \frac{g_*}{6\pi^2} \int_m^\infty dE \frac{(E^2 - m^2)^{3/2}}{\exp[(E - \mu)/T] \pm 1}. \end{aligned} \quad (2.41)$$

Note that there is no integral over d^3x because ρ and P are defined as quantities per unit volume. In the first equality of Eq. (2.41) we have used the fact that the pressure per unit number density of particles is given by $pv/3$ (where v is the speed of the particle), and in the second equality of Eq. (2.41) the relation $v = p/E$ has been used in unit of $c = 1$ (recall that the energy and the momentum in Special Relativity are $E = mc^2/\sqrt{1 - v^2/c^2}$ and $p = mv/\sqrt{1 - v^2/c^2}$, respectively). In the final expressions of Eqs. (2.40) and (2.41) we have adopted the unit $\hbar = 1$. In what follows we consider relativistic and non-relativistic particles separately.

(i) Relativistic species

The relativistic limit corresponds to $T \gg m$, i.e. taking the limit $m \rightarrow 0$ in Eqs. (2.40) and (2.41). For non-degenerate particles ($T \gg \mu$) we obtain

$$\rho = \begin{cases} (\pi^2/30)g_*T^4, & \text{(Bosons)} \\ (7/8)(\pi^2/30)g_*T^4, & \text{(Fermions)} \end{cases} \quad (2.42)$$

$$P = \rho/3, \quad (2.43)$$

where we have used $\int_0^\infty d^3x x^3/(e^x - 1) = \pi^4/15$ and $\int_0^\infty d^3x x^3/(e^x + 1) = 7\pi^4/120$. The result (2.43) shows that the equation of state of relativistic particles without degeneracies is given by $w = 1/3$.

The photons are bosonic species and it is also known that the chemical potential μ for the CMB photons is much smaller than the temperature T ($\mu/T < 9 \times 10^{-5}$ [75]). Since the photon has two spin states ($g_* = 2$), its energy density is

$$\rho_\gamma = \frac{\pi^2}{15} T^4. \quad (2.44)$$

The COBE satellite measured the present temperature of CMB photons to be $T = 2.725 \pm 0.002$ K [12]. On using the conversion $1 \text{ K}^4 = 1.279 \times 10^{-35} \text{ g cm}^{-3}$, the energy density of CMB photons in the present Universe is $\rho_\gamma^{(0)} = 4.641 \times 10^{-34} \text{ g cm}^{-3}$. This corresponds to the density parameter

$$\Omega_\gamma^{(0)} \equiv \frac{8\pi G \rho_\gamma^{(0)}}{3H_0^2} = \frac{\rho_\gamma^{(0)}}{\rho_c^{(0)}} = 2.469 \times 10^{-5} h^{-2}. \quad (2.45)$$

If we take the value $h = 0.72$, then $\Omega_\gamma^{(0)} = 4.763 \times 10^{-5}$. Since the energy density ρ_γ evolves as $\rho_\gamma \propto a^{-4}$ (see Eq. (2.26) with $w = 1/3$), the comparison with Eq. (2.44) gives the relation $T \propto 1/a (= 1 + z)$. Hence the temperature is inversely proportional to the scale factor.

Neutrinos also behave like relativistic particles provided that their masses are small. They are fermionic particles with zero chemical potentials and there are three types of species in standard models (electron neutrino ν_e , muon neutrino ν_μ , and tau neutrino ν_τ). Each species has one spin degree of freedom. Note also that neutrinos have anti-particles (anti-neutrinos). Using Eq. (2.42) in the fermionic case, the energy density of neutrinos, including anti-particles, is given by

$$\rho_\nu = N_{\text{eff}} \frac{7\pi^2}{120} T_\nu^4, \quad (2.46)$$

where N_{eff} is the effective number of neutrino species and T_ν is the temperature of neutrinos. Note that $N_{\text{eff}} = 3$ for standard models of neutrinos, but in the fermionic case we have introduced the effective number N_{eff} in order to allow for other relativistic degrees of freedom.

The Big Bang Nucleosynthesis (BBN) occurred around the energy scale ~ 0.1 MeV to form light elements such as deuterium and helium. The decoupling of neutrinos from the rest of the cosmic plasma, immediately followed by the annihilation of electrons (e^-) and positrons (e^+), occurred earlier than the BBN epoch. The presence of extra relativistic degrees of freedom changes the amount of the light elements predicted by the BBN, which allows to put a bound on N_{eff} . The current standard value is $N_{\text{eff}} = 3.04$ [76], which is slightly larger than 3. The neutrino temperature T_ν is linked to the photon temperature T_γ via the relation $T_\nu/T_\gamma = (4/11)^{1/3}$. This comes from the conservation of entropy before and after the annihilation of electrons and positrons (see the problem 2.2). From Eqs. (2.44) and (2.46), the relation between the neutrino density and the photon density is $\rho_\nu = N_{\text{eff}}(7/8)(4/11)^{4/3} \rho_\gamma$. Hence the present density parameter of radiation, which is the sum of photons and relativistic neutrinos, yields

$$\Omega_r^{(0)} = \frac{\rho_\gamma^{(0)} + \rho_\nu^{(0)}}{\rho_c^{(0)}} = \Omega_\gamma^{(0)}(1 + 0.2271 N_{\text{eff}}), \quad (2.47)$$

where $\Omega_\gamma^{(0)}$ is given in Eq. (2.45). If we take the values $h = 0.72$ and $N_{\text{eff}} = 3.04$, then we obtain $\Omega_r^{(0)} = 8.051 \times 10^{-5}$.

(ii) Non-relativistic matter

In the case of non-relativistic particles ($T \ll m$), Eqs. (2.40) and (2.41) reduce to

$$\rho = g_* m \left(\frac{mT}{2\pi} \right)^{3/2} \exp[-(m - \mu)/T], \quad (2.48)$$

$$P = g_* T \left(\frac{mT}{2\pi} \right)^{3/2} \exp[-(m - \mu)/T] = \frac{T}{m} \rho, \quad (2.49)$$

which are valid for both bosonic and fermionic particles. See problem 2.3 for the derivation of (2.48) and (2.49). Equation (2.49) shows that the pressure P is suppressed relative to the energy density ρ by the factor $T/m \ll 1$. Hence the equation of state for non-relativistic matter is $w \simeq 0$, as expected. The above result shows that the energy density ρ is not described by a function of the temperature T only (unlike the case of photons). Hence we need to measure the density of non-relativistic particles (baryons and dark matter) directly from observations.

Let us consider baryons first. During the BBN epoch the light elements such as deuterium and helium were formed from neutrons and protons. Most of the neutrons decayed to protons (through the β -decay) before the formation of deuterium, while neutrons that did not decay to protons were eventually trapped in helium. If we increase the baryon density, the process of the BBN occurs faster and hence more neutrons remain without decaying to protons. This leads to an increase of the abundance of helium, whereas the abundance of deuterium decreases. Thus the amount of light elements produced during the BBN epoch is sensitive to the baryon density. The abundance of deuterium is known by observing absorption lines in the high-redshift quasars. According to the measurement of distant quasars, Tytler and his collaborators derived the primeval deuterium abundance relative to the hydrogen to be $D/H = (3.0 \pm 0.4) \times 10^{-5}$ [77]. From this bound, Burles, Nollett, and Turner [78] obtained the following constraint on the present density parameter of baryons:

$$\Omega_b^{(0)} h^2 = 0.020 \pm 0.002 \quad (\text{BBN constraint}), \quad (2.50)$$

at the 95% confidence level.

The CMB observations also place tight bounds on the density parameter $\Omega_b^{(0)}$. If we increase the baryon density ρ_b , this leads to a smaller sound speed c_s for the combined fluid system of baryons, photons, and electrons [see Eqs. (4.180) and (4.173)]. Crudely speaking the perturbations in CMB anisotropies with comoving wavenumber k satisfy the equation for the harmonic oscillator, $\frac{d^2 x}{d\eta^2} + k^2 c_s^2 x = 0$, with some corrections [see Eq. (5.16)]. For smaller c_s , the frequency kc_s decreases

so that the height of the first CMB acoustic peak gets larger (because the amplitude of the harmonic oscillator with a smaller spring constant gets larger). One can constrain the amount of the baryon density by using this property. From the WMAP 5-year data combined with SN Ia and BAO data, the constraint on the present density parameter of baryons is [15]

$$\Omega_b^{(0)} h^2 = 0.02267_{-0.00059}^{+0.00058} \quad (\text{WMAP 5 year constraint}), \quad (2.51)$$

at the 68% confidence level. If we take the value $h = 0.72$, then we have $\Omega_b^{(0)} = 0.0437$ for the central value in Eq. (2.51). This means that the baryonic contribution is only 4% in the present Universe.

In addition to baryons, astrophysical observations require the existence of dark matter as another non-relativistic component in the Universe. Since dark matter interacts very weakly with standard model particles, its existence can only be probed by gravitational effects on visible matter. More specifically, if dark matter was non-relativistic at the time it decoupled from photons, it is called Cold Dark Matter (CDM). Alternatively, dark matter that was relativistic at the photon decoupling epoch is called Hot Dark Matter (HDM), whose representative candidate is the neutrino. The present paradigm of structure formation is based on the gravitational clustering of CDM. The baryonic matter alone is not sufficient to lead to structure formation consistent with observations of galaxy clustering [74]. Also the pure HDM model is ruled out as a viable model. This comes from the fact that neutrinos tend to stream out of any overdense region so that the CMB spectrum in the neutrino-dominated Universe has an insufficient power on small scales to be consistent with observations. In the mixed dark matter models of CDM and HDM, the observations limit the amount of hot dark matter to at most a few percent. In the following we shall focus on the pure CDM model (plus dark energy) unless otherwise stated.

The CMB anisotropy data show that the present abundance of dark matter is about 5 times larger than that of baryons. The WMAP 5-year data constrain the density parameter of the CDM to be [15]

$$\Omega_c^{(0)} h^2 = 0.1131 \pm 0.00034 \quad (\text{WMAP 5-year constraint}), \quad (2.52)$$

at the 68% confidence level. For the value $h = 0.72$ we have $\Omega_c^{(0)} = 0.2182$ for the central value in Eq. (2.52).

The origin of dark matter has not been identified yet. There are basically two classes of dark matter – (i) the astrophysical candidates, or (ii) the particle candidates. Examples of class (i) are black holes, neutron stars, and white dwarfs. However, since these originate from baryons, it is not possible to explain all dark matter components without taking into account non-baryonic dark matter.

Class (ii) is of the non-baryonic type. Some examples of this class are axions and Weakly Interacting Massive Particles (WIMPs, including neutralinos). The axion was originally introduced by Peccei and Quinn [79] as a solution to the strong CP problem in quantum chromodynamics (QCD). It has a weak coupling with a small mass $m_a = 10^{-6}$ – 10^{-2} eV. The initial momentum of the axion when it gains a mass through non-perturbative QCD effects is of the order of $p_a = 10^{-9}$ eV $\ll m_a$. Hence the axion can be a good candidate for CDM. The WIMPs are usually motivated by supersymmetric theories. For example, neutralinos are formed as four eigenstates of a mass operator as a result of the mixing of superpartners of Z-bosons, the photon, and the neutral higgs (zino, photino, and higgsino, respectively). The lightest of the four neutralinos turns out to be the lightest supersymmetric particles with typical masses $m_n = 100$ GeV–1 TeV. The lightest neutralinos couple to other particles with the strength characteristic of the weak interaction and hence they can be a good candidate for CDM. Direct or indirect dark matter searches and future LHC experiments will hopefully detect dark matter from space or from high-energy collisions of particles (see Refs. [19, 20, 21] for recent reviews).

(iii) Dark energy

From Eqs. (2.47), (2.51), and (2.52) the sum of the density parameters of radiation, baryons, and dark matter does not exceed 0.3 in the present Universe. Since present observational bounds on the spatial curvature are very strong, $|\Omega_K^{(0)}| \lesssim 0.01$, we still need to identify the remaining 70% of the cosmic matter. This unknown component, called dark energy, is supposed to be responsible for the present cosmic acceleration. The combined data analysis using WMAP, SN Ia, and BAO have provided the following constraint for the present density parameter of dark energy:

$$\Omega_{\text{DE}}^{(0)} = 0.726 \pm 0.015 \quad (\text{WMAP 5-year constraint}). \quad (2.53)$$

In Chapters 5 and 14 we will discuss observational constraints on dark energy in great detail. The theoretical attempts to identify the origin of dark energy will be discussed in Chapters 6–10.

2.4 Cosmic distances

In order to discuss observational constraints on dark energy, it is important to introduce cosmic distances directly related to observations in the FLRW spacetime (2.1). In fact, a large part of the evidence for dark energy comes from measurements of cosmological distances. Setting $r = \sin \chi$ ($K = +1$), $r = \chi$ ($K = 0$), and $r = \sinh \chi$ ($K = -1$) in Eq. (2.2), the 3-dimensional space line-element is expressed as

$$d\sigma^2 = d\chi^2 + (f_K(\chi))^2 (d\theta^2 + \sin^2 \theta d\phi^2), \quad (2.54)$$

where

$$f_K(\chi) = \begin{cases} \sin \chi & (K = +1), \\ \chi & (K = 0), \\ \sinh \chi & (K = -1). \end{cases} \quad (2.55)$$

The function (2.55) can be written in a unified way:

$$f_K(\chi) = \frac{1}{\sqrt{-K}} \sinh \left(\sqrt{-K} \chi \right), \quad (2.56)$$

where the case of the flat universe is recovered by taking the limit $K \rightarrow -0$.

2.4.1 Comoving distance

Let us first compute the comoving distance d_c . The light traveling along the χ direction satisfies the geodesic equation: $ds^2 = -c^2 dt^2 + a^2(t) d\chi^2 = 0$, where we have recovered the speed of light c for clarification. Let us consider the case in which light emitted at time $t = t_1$ with $\chi = \chi_1$ (redshift z) reaches an observer at time $t = t_0$ with $\chi = 0$ (corresponding to $z = 0$). Integrating the equation, $d\chi = -c dt/a(t)$, the comoving distance reads

$$d_c \equiv \chi_1 = \int_0^{\chi_1} d\chi = - \int_{t_0}^{t_1} \frac{c}{a(t)} dt. \quad (2.57)$$

From Eq. (2.28) it follows that $dt = -dz/[H(1+z)]$. Then the comoving distance is given by

$$d_c = \frac{c}{a_0 H_0} \int_0^z \frac{d\tilde{z}}{E(\tilde{z})}, \quad (2.58)$$

where

$$E(z) \equiv H(z)/H_0. \quad (2.59)$$

The function, $\int_0^z d\tilde{z}/E(\tilde{z})$, can be expanded around $z = 0$:

$$\int_0^z \frac{d\tilde{z}}{E(\tilde{z})} = z - \frac{E'(0)}{2} z^2 + \frac{1}{6} \{2E'(0)^2 - E''(0)\} z^3 + \mathcal{O}(z^4), \quad (2.60)$$

where a prime represents a derivative with respect to z . For a redshift z much smaller than unity, the comoving distance is approximately given by

$$d_c \simeq \frac{c}{a_0 H_0} z, \quad \text{for } z \ll 1. \quad (2.61)$$

On using the relation (2.29), we find

$$v \simeq (a_0 H_0) d_c. \quad (2.62)$$

This shows that the recessional velocity v of the object is proportional to d_c with the proportionality constant $a_0 H_0$. For the physical distance $r = a_0 d_c$ we find $r \simeq (c/H_0)z \simeq v/H_0$, which means that Hubble's law (2.32) is satisfied. Hubble's law written as in Eq. (2.32) is valid therefore only in the low-redshift region $z \ll 1$. For $z \gtrsim 1$ the higher-order terms in Eq. (2.60) become important so that Hubble's law is subject to be modified.

2.4.2 Luminosity distance

As we will see in Section 5.2, the luminosity distance d_L is used in SN Ia observations in order to link the supernova luminosity with the expansion rate of the Universe. It is defined by

$$d_L^2 \equiv \frac{L_s}{4\pi \mathcal{F}}, \quad (2.63)$$

where L_s is the absolute luminosity of a source and $\mathcal{F}$ is an observed flux. Note that the observed luminosity L_0 (detected at $\chi = 0$ and $z = 0$) is different from the absolute luminosity L_s of the source (emitted at the comoving distance χ with the redshift z). The flux $\mathcal{F}$ is defined by $\mathcal{F} = L_0/S$, where $S = 4\pi(a_0 f_K(\chi))^2$ is the area of a sphere at $z = 0$. Then the luminosity distance (2.63) yields

$$d_L^2 = (a_0 f_K(\chi))^2 \frac{L_s}{L_0}. \quad (2.64)$$

We need now to derive the ratio L_s/L_0 .

If we write the energy of light emitted at the time-interval Δt_1 to be ΔE_1 , the absolute luminosity is defined by $L_s = \Delta E_1/\Delta t_1$. Similarly the observed luminosity is given by $L_0 = \Delta E_0/\Delta t_0$, where ΔE_0 is the energy of light detected at the time-interval Δt_0 . Since the energy of a photon is inversely proportional to its wavelength λ we have that $\Delta E_1/\Delta E_0 = \lambda_0/\lambda_1 = 1 + z$, where we have used Eq. (2.28). Moreover, the constancy of $c = \lambda/\Delta t$ implies $\lambda_1/\Delta t_1 = \lambda_0/\Delta t_0$, where λ_1 and λ_0 are the wavelength of light at the points of emission and detection respectively. This leads to the relation $\Delta t_0/\Delta t_1 = \lambda_0/\lambda_1 = 1 + z$. Hence we find

$$\frac{L_s}{L_0} = \frac{\Delta E_1}{\Delta E_0} \frac{\Delta t_0}{\Delta t_1} = (1 + z)^2. \quad (2.65)$$

From Eqs. (2.64) and (2.65) the luminosity distance reduces to

$$d_L = a_0 f_K(\chi)(1 + z). \quad (2.66)$$

Recall that the function $f_K(\chi)$ is given in Eq. (2.56) with the comoving distance

$$\chi = d_c = \frac{c}{a_0 H_0} \int_0^z \frac{d\tilde{z}}{E(\tilde{z})}. \quad (2.67)$$

Then d_L can be expressed as

$$d_L = \frac{c(1+z)}{H_0\sqrt{\Omega_K^{(0)}}} \sinh\left(\sqrt{\Omega_K^{(0)}} \int_0^z \frac{d\tilde{z}}{E(\tilde{z})}\right), \quad (2.68)$$

where $\Omega_K^{(0)} = -Kc^2/(a_0H_0)^2$. Note that this definition of $\Omega_K^{(0)}$ is identical to the last one given in Eq. (2.24) in the unit $c = 1$. It is clear that the luminosity distance is directly related to the expansion rate of the Universe.

Expanding the function $\sinh(x)$ in the form $\sinh(x) = x + x^3/6 + \mathcal{O}(x^5)$ and using Eq. (2.60), we find that d_L can be expanded around $z = 0$ as follows:

$$d_L = \frac{c}{H_0} \left[z + \left\{ 1 - \frac{E'(0)}{2} \right\} z^2 + \frac{1}{6} \left\{ 2E'(0)^2 - 3E'(0) - E''(0) + \Omega_K^{(0)} \right\} z^3 + \mathcal{O}(z^4) \right]. \quad (2.69)$$

In the small-redshift region ($z \ll 1$) we have $d_L \simeq cz/H_0$. Using Eq. (2.29) we obtain

$$v \simeq H_0 d_L, \quad \text{for } z \ll 1. \quad (2.70)$$

This shows that Hubble's law holds for the luminosity distance as well.

2.4.3 Angular diameter distance

The angular diameter distance d_A is defined by

$$d_A \equiv \frac{\Delta x}{\Delta\theta}, \quad (2.71)$$

where $\Delta\theta$ is the angle that subtends an object of actual size Δx orthogonal to the line of sight. This distance is often used for the observations of CMB anisotropies.

Since the source lies on the surface of a sphere with radius χ with the observer at the center, the size Δx at time t_1 in the FLRW spacetime (2.1) with (2.54) is given by

$$\Delta x = a(t_1) f_K(\chi) \Delta\theta. \quad (2.72)$$

Hence the diameter distance is

$$d_A = a(t_1) f_K(\chi) = \frac{a_0 f_K(\chi)}{1+z} = \frac{1}{1+z} \frac{c}{H_0\sqrt{\Omega_K^{(0)}}} \sinh\left(\sqrt{\Omega_K^{(0)}} \int_0^z \frac{d\tilde{z}}{E(\tilde{z})}\right), \quad (2.73)$$

where we have used $f_K = c/(a_0 H_0 \sqrt{\Omega_K^{(0)}}) \sinh(\sqrt{\Omega_K^{(0)}} \int_0^z d\tilde{z}/E(\tilde{z}))$ and $z = a_0/a(t_1) - 1$. Comparing Eq. (2.73) with Eq. (2.68), we notice the

following relation

$$d_A = \frac{d_L}{(1+z)^2}. \quad (2.74)$$

This is called reciprocity or duality or Etherington relation [80]. Its validity extends far beyond the FLRW metric: it is valid in fact for any metric as long as flux is conserved.

In the limit $z \ll 1$ all the distances discussed above reduce to the Euclidean distance in the Minkowski spacetime.

2.4.4 Degeneracy of the distance–redshift relation

All the distance definitions given above depend on the cosmological parameters through the integral $\chi = (c/(a_0 H_0)) \int_0^z d\tilde{z}/E(\tilde{z})$ in flat spaces and on $f_K(\chi)$ in curved spaces. It is therefore clear that all the measures of expansion that we can obtain through measurements of distances, from standard candles to the CMB acoustic peaks, will constrain only the cosmological parameters contained in $E(z)$ and only in those particular combinations that appear in χ and in $f_K(\chi)$. If we had distance information only for a given z then all the combinations of cosmological parameters that produce the same $f_K(\chi)$ would be equally acceptable: the constraints would therefore be fully degenerate along lines (or surfaces) of constant $f_K(\chi)$. For instance the closed Λ CDM cosmological model defined by $(\Omega_\Lambda^{(0)}, \Omega_m^{(0)}) = (1, 1)$ and the open model defined by $(\Omega_\Lambda^{(0)}, \Omega_m^{(0)}) = (0.1, 0.6)$ give practically identical distances at $z = 1$. In Fig. 2.1 we plot the lines of constant $f_K(\chi)$ in the plane $(\Omega_\Lambda^{(0)}, \Omega_m^{(0)})$ for redshifts that roughly correspond to a typical distant supernova and to CMB.

If we have information only in a small range of redshifts the degeneracy will be partially broken but still the constraints will appear elongated along the lines of equal distances, as we will see in the next sections. It is only by combining measures at widely different redshifts or by employing indicators other than distances that we may hope to pin down the cosmological parameters.

2.5 The equation of state of dark energy

Let us consider the Universe filled by radiation (density ρ_r and pressure $P_r = \rho_r/3$), non-relativistic matter (density ρ_m and pressure $P_m = 0$), and dark energy (density ρ_{DE} and pressure P_{DE}). Since ρ_r and ρ_m evolve as $\rho_r \propto a^{-4}$ and $\rho_m \propto a^{-3}$, respectively, they can be expressed in the forms

$$\rho_r = \rho_r^{(0)}(a_0/a)^4 = \rho_r^{(0)}(1+z)^4, \quad (2.75)$$

$$\rho_m = \rho_m^{(0)}(a_0/a)^3 = \rho_m^{(0)}(1+z)^3. \quad (2.76)$$

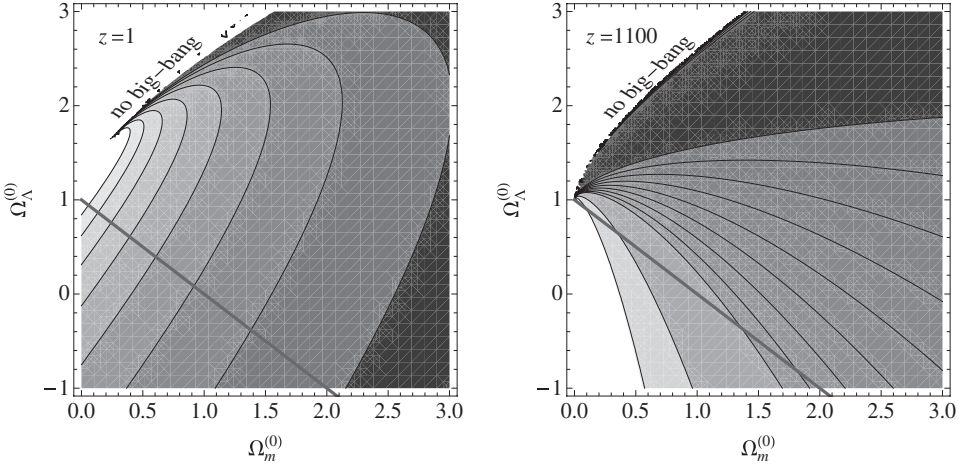


Figure 2.1 Contours of constant $f_K(\chi)$ for $z = 1$ and $z = 1100$. The straight line represents the flat model. In the left panel the contours range from 1400 to $2800 \, h^{-1} \text{ Mpc}$ in steps of 200 , from right to left. In the right panel they range from 300 to $5100 \, h^{-1} \text{ Mpc}$ in steps of 600 , from top to bottom (the last two contours are for $15\,000$ and $30\,000 \, h^{-1} \text{ Mpc}$, respectively).

The redshift z_{eq} that corresponds to the radiation–matter equality ($\rho_r = \rho_m$) is

$$1 + z_{\text{eq}} = \frac{\rho_m^{(0)}}{\rho_r^{(0)}} = \frac{\Omega_m^{(0)}}{\Omega_r^{(0)}}, \quad (2.77)$$

where $\Omega_r^{(0)}$ is given by Eq. (2.47) with Eq. (2.45). The density parameter $\Omega_m^{(0)}$ is the sum of the baryon contribution $\Omega_b^{(0)}$ and the CDM contribution $\Omega_c^{(0)}$, i.e. $\Omega_m^{(0)} = \Omega_b^{(0)} + \Omega_c^{(0)}$. The WMAP 5-year constraints on $\Omega_b^{(0)}$ and $\Omega_c^{(0)}$ are given by Eqs. (2.51) and (2.52), respectively.

For the effective number of neutrino species $N_{\text{eff}} = 3.04$ we obtain

$$1 + z_{\text{eq}} = 2.396 \times 10^4 \Omega_m^{(0)} h^2. \quad (2.78)$$

If we take the value $\Omega_m^{(0)} h^2 = 0.136$, we have $z_{\text{eq}} = 3258$. Note that the CMB decoupling epoch corresponds to $z_{\text{dec}} \simeq 1090$ [15], therefore later than the radiation–matter equality.

Let us consider dark energy with an equation of state $w_{\text{DE}} = P_{\text{DE}}/\rho_{\text{DE}}$, satisfying the continuity equation

$$\dot{\rho}_{\text{DE}} + 3H(\rho_{\text{DE}} + P_{\text{DE}}) = 0. \quad (2.79)$$

Integrating this equation by using the relation $dt = -dz/[H(1+z)]$, we obtain

$$\rho_{\text{DE}} = \rho_{\text{DE}}^{(0)} \exp \left[\int_0^z \frac{3(1+w_{\text{DE}})}{1+\tilde{z}} d\tilde{z} \right], \quad (2.80)$$

which can also be written by introducing an average $\hat{w}_{\text{DE}}$ as

$$\rho_{\text{DE}} = \rho_{\text{DE}}^{(0)} (a/a_0)^{-3(1+\hat{w}_{\text{DE}})}, \quad \hat{w}_{\text{DE}}(z) = \frac{1}{\ln(1+z)} \int_0^z \frac{w_{\text{DE}}(\tilde{z})}{1+\tilde{z}} d\tilde{z}. \quad (2.81)$$

From the Friedmann equation (2.17) we have

$$H^2 = \frac{8\pi G}{3} (\rho_r + \rho_m + \rho_{\text{DE}}) - \frac{K}{a^2}. \quad (2.82)$$

From Eq. (2.82) we see that the present density parameters defined in Eq. (2.24) obey the following relation

$$\Omega_r^{(0)} + \Omega_m^{(0)} + \Omega_{\text{DE}}^{(0)} + \Omega_K^{(0)} = 1. \quad (2.83)$$

Then Eq. (2.82) can be written in the form

$$H^2(z) = H_0^2 \left[\Omega_r^{(0)}(1+z)^4 + \Omega_m^{(0)}(1+z)^3 + \Omega_{\text{DE}}^{(0)} \exp \left\{ \int_0^z \frac{3(1+w_{\text{DE}})}{1+\tilde{z}} d\tilde{z} \right\} + \Omega_K^{(0)}(1+z)^2 \right]. \quad (2.84)$$

Differentiating this equation with respect to z , we find that the equation of state of dark energy can be expressed as

$$w_{\text{DE}}(z) = \frac{(1+z)(E^2(z))' - 3E^2(z) - \Omega_r^{(0)}(1+z)^4 + \Omega_K^{(0)}(1+z)^2}{3 \left[E^2(z) - \Omega_r^{(0)}(1+z)^4 - \Omega_m^{(0)}(1+z)^3 - \Omega_K^{(0)}(1+z)^2 \right]}, \quad (2.85)$$

where $E(z)$ is defined in Eq. (2.59) and a prime represents a derivative with respect to z . From Eq. (2.68) the quantity $E(z)$ can be written in terms of d_L :

$$E^2(z) = \frac{(1+z)^2 \left[c^2(1+z)^2 + \Omega_K^{(0)} H_0^2 d_L(z)^2 \right]}{[(1+z)H_0 d_L'(z) - H_0 d_L(z)]^2}. \quad (2.86)$$

For the flat Universe ($\Omega_K^{(0)} = 0$) this relation reduces to the following simple form

$$E(z) = \frac{c}{H_0} \left[\frac{d}{dz} \left(\frac{d_L(z)}{1+z} \right) \right]^{-1}. \quad (2.87)$$

If the luminosity distance $d_L(z)$ is measured observationally, we can determine the evolution of $E(z)$ from Eq. (2.86) and hence $w_{\text{DE}}(z)$ from Eq. (2.85).

The cosmic expansion history for the redshift $z \lesssim \mathcal{O}(1)$ can be reconstructed from the SN Ia observations. In this regime the energy density of radiation is negligible compared to those of non-relativistic matter and dark energy. The present

observational bound on the cosmic curvature is $-0.0175 < \Omega_K^{(0)} < 0.0085$ [15], showing that the Universe is close to the flat geometry. In the flat Universe with a negligible contribution of radiation, Eq. (2.85) reduces to

$$w_{\text{DE}}(z) = \frac{(1+z)(E^2(z))' - 3E^2(z)}{3 \left[E^2(z) - \Omega_m^{(0)}(1+z)^3 \right]}. \quad (2.88)$$

This relation is often used when we place observational constraints on the equation of state of dark energy. While the effect of the cosmic curvature on the estimate of $w_{\text{DE}}(z)$ can be negligible in the region $z \lesssim 1$, a small uncertainty on the curvature can produce a significant bias in w_{DE} in the high-redshift regime $z \gtrsim 1$ [81, 82]. In such a case we need to use the relation (2.85) rather than (2.88).

It is important to remark however that the basic observable quantity is $E(z)$, not $w_{\text{DE}}(z)$. In fact, $w_{\text{DE}}(z)$ cannot be determined entirely from $E(z)$, i.e. from measurements of the background expansion. From Eq. (2.88) it appears in fact that one needs $\Omega_m^{(0)}$, i.e. the present density of pressureless matter, and this can only be obtained from large-scale structure methods. However one must notice that the density parameter obtained from e.g., the cluster mass estimation does not necessarily coincide with the quantity $\Omega_m^{(0)}$ in Eq. (2.88), since in general clustered matter and pressureless matter do not need to be the same. This is particularly important to notice in coupled dark energy models [17] in which matter acquires an effective pressure through an interaction with dark energy. Of course if $w_{\text{DE}}(z)$ is assumed to be constant or is parametrized in some form, as is usually done, then the knowledge of $E(z)$ at several z 's can fix both the equation of state and $\Omega_m^{(0)}$.

Finally, let us notice that the bound on the equation state of dark energy from the WMAP 5-year data combined with other observational data is $-1.097 < w_{\text{DE}} < -0.858$ at the 95% confidence level [15]. Hence we cannot rule out the possibility that w_{DE} is smaller than -1 . These cases are generally called “phantoms” or “ghosts” [83]. Since $P + \rho < 0$ in this case, Eq. (2.20) shows that ρ increases with time. From Eq. (2.17) the Hubble parameter H grows toward the future. When w_{DE} is constant and smaller than -1 , the solution of the scale factor corresponding to the expanding Universe is given by

$$a \propto (t_{\text{rip}} - t)^{2/(3(1+w_{\text{DE}}))}, \quad (2.89)$$

where t is smaller than the constant t_{rip} . As t approaches t_{rip} , the scale factor goes to infinity. One can easily show that the scalar curvature R and the Hubble parameter H also diverge at $t = t_{\text{rip}}$. The Universe ends up with a finite-time singularity in the (distant!) future (see problem 2.4). This finite-time singularity is called the *big-rip* singularity [84].

2.6 Problems

- 2.1** Derive the continuity equation, $\dot{\rho} + 3H(\rho + P) = 0$, from the continuity equation

$$\nabla_\mu T_0^\mu = 0. \quad (2.90)$$

- 2.2** Let us consider the entropy conservation before and after the annihilation of electrons and positrons when the cosmic temperature was of the order of the electron mass. The entropy density for a particle with density ρ , pressure P , and temperature T is defined by $s \equiv (\rho + P)/T$. Before the annihilation there were photons, neutrinos, anti-neutrinos, electrons, and positrons with the same temperature, whereas after the annihilation there were photons, neutrinos, and anti-neutrinos with different temperatures. By using the entropy conservation as well as the fact that the neutrino temperature scales as $T_\nu \propto a^{-1}$, show that the relation between the photon temperature T_γ and the neutrino temperature T_ν is given by

$$\frac{T_\nu}{T_\gamma} = \left(\frac{4}{11} \right)^{1/3}. \quad (2.91)$$

- 2.3** Derive the density (2.48) and the pressure (2.49) in the non-relativistic limit ($T \ll m$).
2.4 Show that the scalar curvature R diverges at the big-rip singularity by using the solution (2.89). For the dark energy equation of state $w_{\text{DE}} = -1.5$, estimate the time $t_{\text{rip}} - t_0$ by neglecting the contribution of non-relativistic matter.

3

Correlation function and power spectrum

The models of dark energy we will introduce later on are linked to the observations by a precious tool, statistics. Since this world is complicated, we have to average the ups and downs of everyday life to get a sense of the underlying substance. In this chapter we present basic tools of statistics in order to confront dark energy models with observations.

Statistics itself is often divided into *descriptive* statistics, i.e. how to condense the data in a compact and useful way, and *estimation* (or inferential, inductive) statistics, i.e. how to derive information on model parameters. We start here with the statistics needed for cosmological perturbation theory, essentially descriptive statistics such as correlation function and power spectrum, and postpone parameter estimation statistics to Chapter 13.

A note on notation is in order here. When there is no real need, we will not use separate notation for an estimator (say, the correlation function) and its expected value. Similarly, when there is no risk of confusion, we will denote vector quantities like position $\mathbf{x}$ and wavevector $\mathbf{k}$ with unbolded fonts x, k , especially to denote the argument of functions: $\delta(x)$ will in general mean the density contrast at a position $\mathbf{x}$. Finally, when an integration, even a multiple one, is extended to the whole space we simply write $\int dV$ or $\int dx$. When the domain of integration really matters, then it will be specified.

3.1 The correlation function

Our first task is to describe a random distribution of particles in a compact way, to be identified with astrophysical sources (galaxies, quasars, etc.). If there are N points in a volume V , the first interesting descriptor is the average numerical density $\rho_0 = N/V$. Clearly this is utterly insufficient to discriminate among say N points clustered near the same spot and N points evenly distributed across the volume, so we need to find more useful descriptors. Let us focus then on some

small volume dV chosen randomly inside the volume V . Then $\rho_0 dV$ is the average number of particles in the infinitesimal volume. If $dN_{ab} = \langle n_a n_b \rangle$ is the average number of *pairs* in the volumes dV_a and dV_b (i.e. the product of the number of particles in one volume times the number in another volume), separated by r_{ab} , then we define the next important descriptor, the 2-point correlation function $\xi(r_{ab})$, as

$$dN_{ab} = \langle n_a n_b \rangle = \rho_0^2 dV_a dV_b [1 + \xi(r_{ab})]. \quad (3.1)$$

We have implicitly assumed that $r_{ab} > 0$, i.e. the two volumes do not coincide.

The word “average” (and the symbol $\langle \rangle$ we will use to denote it) may have two meanings. One can average by taking many realizations of the distribution, all of them produced in the same way (e.g., by an N -body computer code or by throwing particles at random), selecting in each realization the volumes dV_a , dV_b at the same locations and then averaging the pair number $n_a n_b$. This is the *ensemble* average.

Alternatively one can take the pairs at different spots within the same realization, separated by the same r_{ab} (*sample* average). If the spots are so distant that they are uncorrelated, then we can consider them as coming from different realizations and the two averaging methods coincide. The problem with the approach is that we do not know a priori whether the spots really are uncorrelated until we can compare them with an ensemble of realizations. This issue is particularly acute in astrophysics since we are given a single Universe. The sample correlation function does not in general coincide with the one we would obtain from an ensemble. This problem is of course more important for distributions that are inhomogeneous at very large scales. The estimation of the correlation function on scales smaller than the scale of (approximate) homogeneity will not coincide with the ensemble value. See problem 3.1 for an example. Even in those cases, however, the correlation function is a useful descriptor (although a survey-dependent one) and it makes sense to use it. However here we will always assume that the properties of the sample distribution are a good approximation of the ensemble ones.

If the distribution has been obtained by throwing the N particles at random (i.e. without any preference with respect to the place), then there is no reason for dN_{ab} to depend on the location. Therefore the average number of pairs is exactly equal to the product of the average number of particles in the two volumes, $\langle n_a n_b \rangle = \langle n_a \rangle \langle n_b \rangle = \rho_0^2 dV_a dV_b$, and the correlation ξ vanishes. Conversely, if ξ is non-zero, we say that the particles are correlated. Then the correlation function can be written as a spatial average of the product of the density contrast $\delta(r_a) = n_a/(\rho_0 dV_a) - 1$ at two different points:

$$\xi(r_{ab}) = \frac{dN_{ab}}{\rho_0^2 dV_a dV_b} - 1 = \langle \delta(r_a) \delta(r_b) \rangle, \quad (3.2)$$

where we have used $\langle \delta(r_a) \rangle = \langle \delta(r_b) \rangle = 0$. If this average is taken to be the sample average, then it means we have to average over all possible positions:

$$\xi(\mathbf{r}) = \frac{1}{V} \int \delta(\mathbf{y}) \delta(\mathbf{y} + \mathbf{r}) dV_{\mathbf{y}}. \quad (3.3)$$

A bit of jargon: when $\xi(\mathbf{r})$ depends only on the separation $\mathbf{r}$ and not on $\mathbf{r}_a$ and $\mathbf{r}_b$ individually, the system is said to be *statistically homogeneous* (i.e. it possesses the same statistical properties everywhere), although strictly speaking for this property to be true all the higher-order statistics should also be independent of location. If moreover the ensemble average coincides with the sample average, then the system is said to be *ergodic*. However the latter term refers historically to time processes, not to spatial ones. The term most often used in astrophysics is that the system is a *fair sample* of the Universe.

In practice it is easier to derive the correlation function as the average density of particles at a distance r from another particle, i.e. by choosing the volume dV_a so that $\rho_0 dV_a = 1$. Then the number of pairs is given by the number of particles in the volume dV_b :

$$dN_b = \rho_0 dV_b [1 + \xi(r_b)]. \quad (3.4)$$

Operationally therefore one evaluates the correlation function as follows:

$$\xi(r) = \frac{dN(r)}{\rho_0 dV} - 1 = \frac{\langle \rho_c \rangle}{\rho_0} - 1, \quad (3.5)$$

i.e. as the average number of particles at distance r from any given particle (or number of neighbors), divided by the expected number of particles at the same distance in a uniform distribution, minus 1 (sometimes this is called *conditional density contrast*). In a finite volume with $N = \rho_0 V$ particles one clearly has an integral constraint on $\xi(r)$ due to the fact that the average density is calculated within the volume itself:

$$\int \xi(r) dV = \frac{1}{\rho_0} \int \frac{dN}{dV} dV - V = \frac{N}{\rho_0} - V = 0. \quad (3.6)$$

If the correlation $\xi(r)$ is positive, there are more particles than in a uniform distribution. In this case the distribution is said to be positively clustered. Quite often one is interested only in the dependence on the modulus r , so the volume at distance r is chosen as a shell around each particle. One could generalize this definition by introducing the anisotropic correlation function as the number of pairs in volumes separated by the vector $\mathbf{r}$. This is useful whenever there is some reason to suspect that the distribution is indeed anisotropic, as when there is a significant distortion along the line of sight due to the redshift.

The estimator (3.5) requires the knowledge of the number density ρ_c inside a shell of thickness dr at distance r from every particle. In other words, it requires the estimation of the density in every shell. In practice, a direct estimation of the shell density is difficult because of the complicated boundary and selection procedure that a real survey often has. The simplest way to measure ξ is to compare the real catalog to an artificial random catalog with exactly the same boundaries and the same selection function. Then the estimator can be written as

$$\xi = \frac{DD}{DR} - 1, \quad (3.7)$$

where DD means the number of galaxies at distance r counted by an observer centered on a real galaxy (data D). This is divided by the number of galaxies DR at the same distance but in the random catalog (if to reduce the scatter the random catalog contains α times more galaxies than the real one then DR has to be divided by α). In other words, instead of calculating the volume of the shell (which is a difficult task in realistic cases), we estimate ξ by counting the galaxies in the Monte Carlo realization. In this way all possible boundaries and selection function can be easily mimicked in the random catalog, which will affect DD and DR in the same way. Following these considerations, similar estimators with numerically more robust properties have been proposed [85].

3.2 The n -point correlation function

The average number density ρ_0 can be called a one-point estimator of a random field, since it is estimated by averaging an ensemble over the same location. The correlation function we have just seen is a two-point estimator because it requires averaging over two small volumes. It is clear that one can extend these definitions to higher orders. For instance, the three-point function is defined as

$$\zeta_{abc}(r_a, r_b, r_c) = \langle \delta(r_a) \delta(r_b) \delta(r_c) \rangle. \quad (3.8)$$

In terms of the counts in infinitesimal cells, we can write

$$\begin{aligned} \zeta_{abc}(r_a, r_b, r_c) &= \left\langle \left(\frac{n_a}{\rho_0 dV_a} - 1 \right) \left(\frac{n_b}{\rho_0 dV_b} - 1 \right) \left(\frac{n_c}{\rho_0 dV_c} - 1 \right) \right\rangle \\ &= \frac{\langle n_a n_b n_c \rangle}{\rho_0^3 dV_a dV_b dV_c} - \xi_{ab} - \xi_{bc} - \xi_{ac} - 1, \end{aligned} \quad (3.9)$$

where $\xi_{ij} \equiv \xi(r_{ij})$. We then obtain the following useful relation

$$\langle n_a n_b n_c \rangle = \rho_0^3 dV_a dV_b dV_c (1 + \xi_{ab} + \xi_{bc} + \xi_{ac} + \zeta_{abc}), \quad (3.10)$$

where ς_{abc} is called the “disconnected” part of the third-order correlation function. A random field is said to be Gaussian when $\varsigma_{abc} = 0$ (and the same for all disconnected higher-order moments). In this case the two-point correlation function (or its Fourier version, the power spectrum, see the next section) completely describes the statistical properties of the field.

3.3 The power spectrum

As we will see in the next chapters, in particular in Chapter 4, the linear perturbation variables contain important physics, both for dark energy and cosmology in general. A convenient way to study perturbation variables is to decompose fluctuations into orthonormal modes because at the linear level they evolve independently. Since by definition the average of a perturbation variable is zero, the simplest non-trivial statistics corresponds to a quadratic function of the variables. In Fourier space, any real quadratic function of a perturbation variable is called a *power spectrum*. Examples are

$$P_\delta(k) = A|\delta_k|^2, \quad (3.11)$$

$$P_\Phi(k) = B|\Phi_k|^2, \quad (3.12)$$

where δ_k and Φ_k are the Fourier coefficients of the density contrast and the gravitational potential, respectively, with A and B being some constants. If the quadratic form is composed of two different variables, e.g., $|\delta_k\theta_k|$, then we have a cross-correlation power spectrum. If we average over directions, the power spectrum will depend only on the modulus k . The power spectrum is by far the most common descriptor of clustering in the linear and mildly non-linear regime and plays a central role in cosmology and in this book. As we will show soon, the power spectrum is the Fourier-space version of the correlation spectrum.

Unless otherwise specified, the normalization convention for the 3-dimensional (3D) Fourier transformation is

$$f(\mathbf{x}) = \frac{V}{(2\pi)^3} \int f_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}} d^3k, \quad (3.13)$$

$$f_{\mathbf{k}} = \frac{1}{V} \int f(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} d^3x. \quad (3.14)$$

With these conventions, $f(\mathbf{x})$ and $f_{\mathbf{k}}$ have the same dimensions. Remember however that for the normalization of the Fourier transformation the only important property is that the product of the pre-integral factors in 3 dimensions amounts to $(2\pi)^{-3}$. For all theoretical manipulations the factor of V can well be replaced by unity and often for convenience we will do so. Dirac’s delta function $\delta_D(\mathbf{x})$ is

defined as

$$\delta_D(\mathbf{x}) = (2\pi)^{-3} \int e^{i\mathbf{k}\cdot\mathbf{x}} d^3k . \quad (3.15)$$

Analogous definition holds for Dirac's function in Fourier space (which is not the Fourier transform of $\delta_D(\mathbf{x})$)

$$\delta_D(\mathbf{k}) = (2\pi)^{-3} \int e^{i\mathbf{k}\cdot\mathbf{x}} d^3x , \quad (3.16)$$

and their normalization is such that

$$\int \delta_D(\mathbf{k}) d^3k = \int \delta_D(\mathbf{x}) d^3x = 1 . \quad (3.17)$$

For the density contrast of a density field $\delta(\mathbf{x})$, the Fourier transform is

$$\delta_k = \frac{1}{V} \int \delta(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} dV . \quad (3.18)$$

The power spectrum is defined as

$$P(\mathbf{k}) = V |\delta_k|^2 = V \delta_k \delta_k^* . \quad (3.19)$$

Notice that the power spectrum has the dimension of a volume. It follows that

$$P(\mathbf{k}) = \frac{1}{V} \int \delta(\mathbf{x}) \delta(\mathbf{y}) e^{-i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})} dV_x dV_y . \quad (3.20)$$

Setting $\mathbf{r} = \mathbf{x} - \mathbf{y}$, the spectrum (3.20) reduces to

$$P(\mathbf{k}) = \int \xi(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}} dV , \quad (3.21)$$

where $\xi(\mathbf{r})$ is defined in Eq. (3.3):

$$\xi(\mathbf{r}) = \langle \delta(\mathbf{y} + \mathbf{r}) \delta(\mathbf{y}) \rangle = \frac{1}{V} \int \delta(\mathbf{y} + \mathbf{r}) \delta(\mathbf{y}) dV_y . \quad (3.22)$$

Therefore, the power spectrum is the Fourier transform of the correlation function (Wiener–Khinchin theorem). The converse property is

$$\xi(\mathbf{r}) = (2\pi)^{-3} \int P(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{r}} d^3k . \quad (3.23)$$

Notice that here and in the following the Fourier volume factor is not included, as in most literature. Assuming spatial isotropy, i.e. that the correlation function depends only on the modulus $r = |\mathbf{r}|$, the spectrum depends only on $k = |\mathbf{k}|$:

$$P(k) = \int \xi(r) r^2 dr \int_0^\pi e^{-ikr \cos \theta} \sin \theta d\theta \int_0^{2\pi} d\phi = 4\pi \int \xi(r) \frac{\sin kr}{kr} r^2 dr . \quad (3.24)$$

The integral constraint (3.6) implies $P(0) = 0$ in k -space: in any finite survey, the spectrum vanishes at large scales.

A more general definition of the power spectrum can also be given, but this time we have to think in terms of ensemble averages rather than volume averages. Consider in fact the ensemble average of $V\delta_k\delta_{k'}^*$:

$$\begin{aligned} V\langle\delta_k\delta_{k'}^*\rangle &= \frac{1}{V} \int \langle\delta(\mathbf{x})\delta(\mathbf{y})\rangle e^{-i\mathbf{k}\cdot\mathbf{x}+i\mathbf{k}'\cdot\mathbf{y}} dV_x dV_y \\ &= \frac{1}{V} \int \langle\delta(\mathbf{y})\delta(\mathbf{y}+\mathbf{r})\rangle e^{-i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{y}-i\mathbf{k}\cdot\mathbf{r}} dV_r dV_y. \end{aligned} \quad (3.25)$$

In order to perform ensemble averages, one has to think of fixing a position and taking the average over the ensemble of realizations. Then the average can enter the integration and acts only over the random variables δ . Then we obtain

$$\begin{aligned} V\langle\delta_k\delta_{k'}^*\rangle &= \frac{1}{V} \int \xi(\mathbf{r}) e^{-i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{y}-i\mathbf{k}\cdot\mathbf{r}} dV_r dV_y \\ &= \frac{1}{V} \int e^{-i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{y}} dV_y \int e^{-i\mathbf{k}\cdot\mathbf{r}} \xi(\mathbf{r}) dV_r \\ &= \frac{(2\pi)^3}{V} P(\mathbf{k}) \delta_D(\mathbf{k}-\mathbf{k}'). \end{aligned} \quad (3.26)$$

The definition (3.26) shows that modes at different wavelengths are uncorrelated if the field is statistically homogeneous (that is, if ξ does not depend on the position $\mathbf{y}$ but only on the separation $\mathbf{r}$). Since $\delta(\mathbf{x})$ is a real function, $\delta(\mathbf{k}) = \delta^*(-\mathbf{k})$ and therefore the form $V\langle\delta_k\delta_{k'}\rangle = (2\pi)^3 V^{-1} P(\mathbf{k}) \delta_D(\mathbf{k}+\mathbf{k}')$ can also be employed.

In our cosmological applications we often wish to study a continuous underlying field, for instance the gravitational potential or the dark matter density contrast δ_m . We think of the galaxies as a mere discrete tracer of this field, as sand particles trace the wind in a sandstorm. The galaxies form therefore a discrete sampling of the underlying field. The only way to get information on the underlying field is to study this discrete sampling. Although we may assume by and large that “more galaxies” mean also “larger dark matter density contrast,” i.e. that the average of the galaxy density contrast δ_g is everywhere equal to δ_m , we need to take into account possible departure from the ideal case.

The simplest case is when the number of galaxies in a given location follows a Poissonian distribution with the expected mean value δ_m . This would be the case if the galaxy formation process were indeed a Poissonian process: the outcome of many independent factors each with the same, small probability to occur. We do not know yet if this is an accurate description of what happens but at least it is a reasonable starting hypothesis.

To investigate the discreteness, we assume as field a collection of N particles of dimensionless unitary masses at positions $\mathbf{x}_i$, in a volume V . In the following we will make use of the window function $W(\mathbf{x})$, a function that expresses the way in which the particles are selected. A typical selection procedure is to take all particles within a given region, and no particles elsewhere. In this case, the function will be a constant inside the survey, and zero outside (top-hat window function). We will always consider such a kind of window function in the following, and normalize it such that

$$\int W(\mathbf{x}) dV = 1. \quad (3.27)$$

With this normalization, $W(\mathbf{x}) = 1/V$ inside the survey. The density contrast field we have in a specific sample is therefore the field times the window function (times the sample volume V because of the way we have normalized W):

$$\delta_s = \delta(\mathbf{x}) V W(\mathbf{x}). \quad (3.28)$$

Let us now express the field as a sum of Dirac functions $\rho(\mathbf{x}) = \sum_i \delta_D(\mathbf{x} - \mathbf{x}_i)$ so that

$$\delta_s(\mathbf{x}) = \left(\frac{\rho(\mathbf{x})}{\rho_0} - 1 \right) V W(\mathbf{x}) = \frac{V}{N} \sum_i w_i \delta_D(\mathbf{x} - \mathbf{x}_i) - V W(\mathbf{x}), \quad (3.29)$$

where $w_i = V W(\mathbf{x}_i)$ and as usual $\rho_0 = N/V$. The Fourier transform is

$$\delta_k = \frac{1}{V} \int \left(\frac{V}{N} \sum_i w_i \delta_D(\mathbf{x} - \mathbf{x}_i) - V W(\mathbf{x}) \right) e^{-ik \cdot \mathbf{x}} dV = \frac{1}{N} \sum_i w_i e^{-ik \cdot \mathbf{x}_i} - W_k, \quad (3.30)$$

where we have introduced the k -space window function

$$W_k = \int W(\mathbf{x}) e^{-ik \cdot \mathbf{x}} dV, \quad (3.31)$$

normalized such that $W_0 = 1$. The spherical top-hat function corresponds to

$$\begin{aligned} W(x) &= 1/V, & \text{inside a spherical volume } V \text{ of radius } R, \\ W(x) &= 0, & \text{outside.} \end{aligned} \quad (3.32)$$

We then have

$$W_k = V^{-1} \int_V e^{-ik \cdot \mathbf{x}} dV = \frac{3}{R^3} \int_0^R \frac{r \sin kr}{k} dr = \frac{3(\sin kR - kR \cos kR)}{(kR)^3}, \quad (3.33)$$

where in the second equality we have integrated out the angular part as we did to derive Eq. (3.24). Notice that the function declines rapidly as $k \rightarrow \pi/R$. Now

squaring and averaging $\delta_{\mathbf{k}}$ in Eq. (3.30) by separating the $i = j$ terms from the others, we have

$$\langle \Delta^2(\mathbf{k}) \rangle \equiv V \langle \delta_{\mathbf{k}} \delta_{\mathbf{k}}^* \rangle = P(\mathbf{k}) + P_n, \quad (3.34)$$

where the “true” spectrum $P(\mathbf{k})$ and the pure noise spectrum are given, respectively, by

$$P(\mathbf{k}) = \frac{V}{N^2} \sum_{i \neq j} \langle w_i w_j \rangle e^{-i\mathbf{k} \cdot (\mathbf{x}_i - \mathbf{x}_j)} - V W_k^2, \quad (3.35)$$

$$P_n = \frac{V}{N^2} \sum_i w_i^2 = \frac{V}{N}, \quad (3.36)$$

where the last equality holds if w_i equals 0 or 1. In order to derive Eq. (3.34) we have used the relation $\langle \delta_{-\mathbf{k}} \rangle = 0$

$$\left\langle \frac{1}{N} \sum_i w_i e^{-i\mathbf{k} \cdot \mathbf{x}_i} \right\rangle = W_k. \quad (3.37)$$

The noise spectrum, negligible only for large densities, $\rho_0 = N/V \rightarrow \infty$, is the power spectrum of a distribution with no intrinsic correlation, i.e. obtained by throwing the particles at random. More exactly, it is the power spectrum of a Poissonian distribution. Since the galaxy distributions are often sparse, the noise is not always negligible and has to be subtracted from the estimate if we want to detect the underlying correlation. Therefore the *estimator* of the “true” power spectrum $P(\mathbf{k})$ can be taken as

$$\hat{P}(\mathbf{k}) = \Delta^2(\mathbf{k}) - P_n. \quad (3.38)$$

As for the correlation function, the power spectrum does not characterize a distribution completely, unless we know the distribution has some specific property, e.g., Gaussian, or Poisson, etc. In particular, if we assume the fluctuations to be Gaussian, we can derive the variance of the power spectrum, defined as

$$\sigma_P^2 \equiv \langle [\hat{P}(\mathbf{k}) - P(\mathbf{k})]^2 \rangle = \langle \Delta^4(\mathbf{k}) \rangle - \langle \Delta^2(\mathbf{k}) \rangle^2, \quad (3.39)$$

where $P(\mathbf{k}) \equiv \langle \hat{P}(\mathbf{k}) \rangle = \langle \Delta^2(\mathbf{k}) \rangle - P_n$. To evaluate $\Delta^4(\mathbf{k})$ we proceed as follows. Neglecting the window term (i.e. in the limit of a large volume) the continuous fluctuation field in Fourier space can be written in a compact way as

$$\delta_{\mathbf{k}} = \sum_i \frac{g_i}{V^{1/2}} e^{-i\mathbf{k} \cdot \mathbf{r}_i}, \quad (3.40)$$

where by assumption g_i are Gaussian random variables (the volume factor is only for convenience). The power spectrum is $\Delta^2(\mathbf{k}) = V \langle \delta_{\mathbf{k}} \delta_{\mathbf{k}}^* \rangle = \sum_i g_i^2$. We want to

evaluate

$$\langle \Delta^2(\mathbf{k}) \Delta^2(\mathbf{k}') \rangle = V^2 \langle \delta_{\mathbf{k}} \delta_{\mathbf{k}'}^* \delta_{\mathbf{k}'}^* \delta_{\mathbf{k}}^* \rangle = \sum_{ijmn} \langle g_i g_j g_m g_n \rangle e^{-i(\mathbf{k} \cdot \mathbf{r}_i - \mathbf{k} \cdot \mathbf{r}_j + \mathbf{k}' \cdot \mathbf{r}_m - \mathbf{k}' \cdot \mathbf{r}_n)}. \quad (3.41)$$

Now, the oscillating terms in the sum are negligible except when $i = j = m = n$ or when the indices are equal in pairs. Then we can write

$$\langle \Delta^2(\mathbf{k}) \Delta^2(\mathbf{k}') \rangle = \sum_i \langle g_i^4 \rangle + \sum_i \sum_{j \neq i} \langle g_i^2 \rangle \langle g_j^2 \rangle \left[1 + e^{-i(\mathbf{k} + \mathbf{k}') \cdot (\mathbf{r}_i - \mathbf{r}_j)} + e^{-i(\mathbf{k} - \mathbf{k}') \cdot (\mathbf{r}_i - \mathbf{r}_j)} \right]. \quad (3.42)$$

Since g_i is a Gaussian variable, all odd moments are zero and even moments can be written in terms of the variance; in particular, we have $\langle g_i^4 \rangle = 3 \langle g_i^2 \rangle^2$. This is exactly what is needed to supply the sum for $j \neq i$ with the missing terms $j = i$. Moreover, the oscillating exponential terms average out except for $\mathbf{k}$ close to $\mathbf{k}'$ (or equivalently $-\mathbf{k}'$). Then for $\mathbf{k} \approx \mathbf{k}'$ we can neglect the $\mathbf{k} + \mathbf{k}'$ term since its faster oscillations are averaged over and therefore suppressed relative to the $\mathbf{k} - \mathbf{k}'$ term. Then we have

$$\langle \Delta^2(\mathbf{k}) \Delta^2(\mathbf{k}') \rangle = \sum_i \sum_j \langle g_i^2 \rangle \langle g_j^2 \rangle \left[1 + e^{-i(\mathbf{k} - \mathbf{k}') \cdot (\mathbf{r}_i - \mathbf{r}_j)} \right], \quad (3.43)$$

which amounts to

$$\langle \Delta^2(\mathbf{k}) \Delta^2(\mathbf{k}') \rangle = \langle \Delta^2(\mathbf{k}) \rangle \langle \Delta^2(\mathbf{k}') \rangle + V^2 \langle \delta_{\mathbf{k}} \delta_{\mathbf{k}'}^* \rangle^2. \quad (3.44)$$

From Eqs. (3.39) and (3.34) it follows that

$$\sigma_P^2(\mathbf{k}) = V^2 \langle \delta_{\mathbf{k}} \delta_{\mathbf{k}}^* \rangle^2 = (P(\mathbf{k}) + P_n)^2. \quad (3.45)$$

We obtain finally for the fractional variance of the power spectrum with top-hat filtering a very simple and useful result

$$\frac{\sigma_P^2(\mathbf{k})}{P^2(\mathbf{k})} = \left(1 + \frac{1}{nP(\mathbf{k})} \right)^2, \quad (3.46)$$

where $n = N/V$ is the number density. This tells us that, if the fluctuations are Gaussian, the error of the root mean square (rms) on the power spectrum is of the order of the power spectrum itself (including the shot noise).

In general we consider the shell-averaged spectrum, i.e. the spectrum for all modes whose wavenumber modulus k lies within the shell Δk of volume V_k :

$$P(k) = \frac{1}{V_k} \int_{\Delta k} P(\mathbf{k}') d^3 k'. \quad (3.47)$$

If the survey has a volume $V_s = L^3$, the lowest wavenumber we can safely construct is $k_{\min} = 2\pi/L$. Then the number of *independent* k -modes in a volume V_k is

$$N_k = \frac{V_k}{k_{\min}^3} = \frac{V_k V_s}{(2\pi)^3}. \quad (3.48)$$

Therefore the error on the shell-averaged spectrum $P(k)$ is reduced by the factor $1/N_k$ and we obtain

$$\frac{\sigma_P^2(k)}{P^2(k)} \simeq \frac{(2\pi)^3}{V_k V_s} \left(1 + \frac{1}{nP}\right)^2. \quad (3.49)$$

Another way of looking at this equation is to say that the effective k -volume resolution $k_{\min}^3$ degrades due to the shot noise to $k_{\min}^3(1 + 1/nP)^2$, so that there are effectively less independent k -volumes to average over.

A more complete derivation including an arbitrary density field and selection function is given in Ref. [86]. The general formula is

$$\frac{\sigma_P^2(k)}{P^2(k)} = \frac{(2\pi)^3 \int_{V_s} d^3r n^4 w^4 [1 + 1/(nP(k))]^2}{V_k [\int_{V_s} d^3r n^2 w^2]^2}, \quad (3.50)$$

where $n = n(r)$ is the average density and $w = w(r)$ is an arbitrary weight that might be modulated to minimize the variance itself. If n and w are constants, this expression reduces to Eq. (3.49).

Finally one should be aware of the problem hidden in the result above. In order to evaluate the variance σ_P , the spectrum itself should be known. In practice, this means that one has to guess a spectrum before its variance can be evaluated. More consistently, in the likelihood method procedure of Chapter 13, we will take into account the fact that the variance depends on the model itself.

3.4 From the power spectrum to the moments

Since the power spectrum is often the basic outcome of structure formation theories, it is convenient to express all the other quantities in terms of it. Here we find the relation between the power spectrum and the moments of the counts in random cells.

Consider a finite cell. Divide it into infinitesimal cells with counts n_i either zero or unity, so that for any positive power m we have

$$\sum_i \langle n_i^m \rangle = N_0, \quad (3.51)$$

where $N_0 = \rho_0 V$ is the count average. We have by definition of ξ for $i \neq j$:

$$\langle n_i n_j \rangle = \rho_0^2 (1 + \xi_{ij}) dV_i dV_j. \quad (3.52)$$

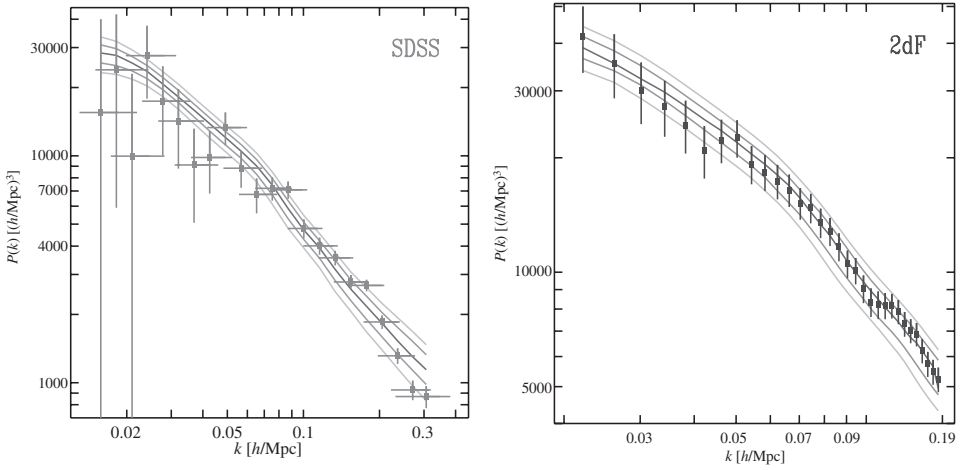


Figure 3.1 (Left) The power spectrum inferred from the SDSS galaxy power spectrum (data points from [65]) is compared to the predicted power spectrum based on the range of parameters consistent with the WMAP parameters (continuous lines). The galaxy power spectrum is normalized by weak lensing measurements [87]. (Right) The predicted power spectrum is compared to the mass power spectrum inferred from the 2-degree Field (2dF) Galaxy Redshift Survey [88]. From Ref. [14].

The count in the cell is $N = \sum_i n_i$. The variance (the second-order moment) is then given by

$$M_2 = N_0^{-2} \langle (\Delta N)^2 \rangle = N_0^{-2} (\langle N^2 \rangle - N_0^2), \quad (3.53)$$

where $\Delta N = N - N_0$. The expectation value $\langle N^2 \rangle$ is

$$\langle N^2 \rangle = \left\langle \sum_i n_i \sum_j n_j \right\rangle = \sum_i \langle n_i^2 \rangle + \sum_{i \neq j} \langle n_i n_j \rangle = N_0 + N_0^2 \int W_i W_j (1 + \xi_{ij}) dV_i dV_j, \quad (3.54)$$

where W_i is the window function and $\xi_{ij} \equiv \xi(|\mathbf{r}_i - \mathbf{r}_j|)$. We define the following integral (by definition $\int W_i dV = 1$ for any window function)

$$\sigma^2 = \int W_1 W_2 \xi_{12} dV_1 dV_2. \quad (3.55)$$

Then Eq. (3.54) reduces to

$$\langle N^2 \rangle = N_0 + N_0^2 (1 + \sigma^2). \quad (3.56)$$

Since $\xi_{12} = (2\pi)^{-3} \int P(\mathbf{k}) e^{i\mathbf{k} \cdot (\mathbf{r}_1 - \mathbf{r}_2)} d^3 k$ from Eq. (3.23), we have

$$\sigma^2 = (2\pi)^{-3} \int P(\mathbf{k}) e^{i\mathbf{k} \cdot (\mathbf{r}_1 - \mathbf{r}_2)} W_1 W_2 d^3 k d^3 r_1 d^3 r_2, \quad (3.57)$$

where the integral $dV_1 dV_2$ in Eq. (3.55) is replaced by $d^3r_1 d^3r_2$. For *spherical cells* of radius R , integrating over the angles, this reduces to

$$\sigma_R^2 = \frac{1}{2\pi^2} \int P(k) W_R^2(k) k^2 dk, \quad (3.58)$$

where we have used $W_R^2(k) = \int e^{ik \cdot r_1} W_1 d^3r_1 \int e^{-ik \cdot r_2} W_2 d^3r_2$.

If the cells have a radius of $8 h^{-1}$ Mpc, it turns out that σ_R is close to unity. Conventionally the normalization of the power spectrum is therefore given by quoting σ_8 .

Substituting Eq. (3.56) into Eq. (3.53), we find that the second-order moment M_2 has the following relation with the power spectrum amplitude σ^2 :

$$M_2 = N_0^{-1} + \sigma^2. \quad (3.59)$$

The first and the second terms correspond to the noise and the count variance in the continuous limit, respectively.

For the third-order moment we proceed in a similar fashion and we obtain (see problem 3.3)

$$M_3 = N_0^{-2} + \int W_i W_j W_k S_{ijk} dV_i dV_j dV_k. \quad (3.60)$$

Of course similar relations can be found at any order. Non-zero higher-order moments are useful to quantify the deviation from Gaussianity of the matter and galaxy distribution.

3.5 Problems

- 3.1** Find the 3-dimensional correlation function of points distributed randomly on the equatorial plane of a sphere.
- 3.2** Using Eq. (3.23), find the isotropic correlation function of a distribution whose isotropic power spectrum is Gaussian with $k \geq 0$ peaked at $k = 0$ with variance σ_k^2 .
- 3.3** Derive Eq. (3.60).

4

Basics of cosmological perturbation theory

In this chapter we present the basics of linear perturbations in cosmology. After a general introduction of cosmological perturbation theory, we work out several cases: (i) single pressureless perfect fluid, (ii) single general perfect fluid, and (iii) two fluids: matter and radiation. We also discuss a number of topics such as the velocity field, the redshift distribution, Boltzmann equations, the matter power spectrum, and the perturbed photon propagation. These provide us with important tools when we confront dark energy models with observations of the cosmic microwave background (CMB) and large-scale structure (LSS).

In this chapter the treatment and notation are fairly standard and the topic is covered in most modern textbooks on cosmology. Readers familiar with cosmological perturbation theory may skip this chapter.

4.1 Perturbing General Relativity

In Chapter 2 we have outlined the cosmic expansion history in the homogeneous and isotropic FLRW background. However, our Universe is far richer than this simple picture. A metric that deviates from the FLRW spacetime can be written as the sum of an unperturbed FLRW part plus something else, that we can generally call “perturbed” metric. If the perturbed part is assumed to be small, in a sense to be defined later, then this splitting of the full metric into a background part and a perturbed one leads to extremely useful results. As we all know, physics is to a large extent described by a Taylor expansion to some low order, and cosmology is not an exception.

To perturb the relativistic equations one must first of all perturb the metric, writing at first order

$$g_{\mu\nu} = g_{\mu\nu}^{(0)} + \delta g_{\mu\nu} , \quad (4.1)$$

where all the entries in the perturbed metric $\delta g_{\mu\nu}$ have to be small with respect to the 0-th order part. In this chapter we write the metric directly in terms of the conformal time $\eta = \int a^{-1} dt$. We consider cosmological perturbations about the flat FLRW metric given by

$$ds^2 = g_{\mu\nu}^{(0)} dx^\mu dx^\nu = a^2(-d\eta^2 + \delta_{ij} dx^i dx^j). \quad (4.2)$$

We will also use the conformal Hubble quantity

$$\mathcal{H} \equiv \frac{1}{a} \frac{da}{d\eta} = Ha. \quad (4.3)$$

In General Relativity the field equations are invariant under a general coordinate change. This means that the split between a background metric and a perturbed one is not unique. However, although it is often a great simplification to choose some special coordinate frame, it would be very confusing if we change in the process also the unperturbed (or background) metric. We would like for instance to keep the FLRW metric as “the” background whenever we make a general transformation. Therefore we select a class of infinitesimal transformations that leaves $g_{\mu\nu}^{(0)}$ as it is, while the perturbed metric $\delta g_{\mu\nu}$ is subject to change. Following general physics usage, this class of transformations is called *gauge transformations*.

In the unperturbed Universe, we have already defined comoving coordinates in such a way that the matter particles expanding with the Universe remain at fixed (comoving) coordinates. When perturbations are added, we can either use the same coordinates, or set up a new set of coordinates that free-fall with the particles in the perturbed gravitational field, or even adopt a totally different frame not related to matter particles. That is, for instance, we can choose to attach the observers to the points in the unperturbed frame or to the perturbed particles. In the former case, to be called the *Newtonian* or *longitudinal gauge*, the observers will detect a velocity field of particles falling into the clumps of matter and will measure a gravitational potential. This choice is in fact the most intuitive one and reduces easily to the Newtonian case. On the other hand, when the wavelengths of perturbations are larger than the horizon, to attach observers to an invisible background is not a convenient choice. In the second case, called the *comoving proper-time gauge*, the observers are attached instead to the free-falling particles, so they do not see any velocity field (unless there are other non-gravitational forces, like pressure gradients) and, being always free falling, do not measure a gravitational potential. This gauge, therefore, does not have a straightforward Newtonian limit. One can also define quantities that are gauge-invariant, but this does not necessarily simplify the understanding and often leads to cumbersome mathematics.

Although a gauge transformation is expressed as a change of coordinates, it is important to observe that (unlike ordinary coordinate transformations) it does not

link different observers in the same spacetime but it links two different spacetimes, the background and the perturbed one, seen by the same observer. This is the reason, for instance, why scalar quantities *change* under a gauge transformation but not under a coordinate transformation.

For a detailed discussion of gauge choices, see Refs. [70, 89, 90, 91, 92, 93, 94].

4.2 The Newtonian gauge

Let us start then with the Newtonian gauge. The most general perturbed metric can be written schematically as $g_{\mu\nu} = g_{\mu\nu}^{(0)} + \delta g_{\mu\nu}$, where

$$\delta g_{\mu\nu} = a^2 \begin{pmatrix} -2\Psi & w_i \\ w_i & 2\Phi\delta_{ij} + h_{ij} \end{pmatrix}. \quad (4.4)$$

Here Ψ and Φ are spatial scalars, w_i is a 3-vector, and h_{ij} is a traceless 3-tensor. All the perturbation quantities (Ψ , Φ , w_i , etc.) depend on space and time although we will not necessarily indicate this explicitly for simplicity of notation. For instance, it is easy to show that g_{00} is a spatial scalar. If $\tilde{x}^\mu = f(x^\mu)$ is a general transformation, the metric tensor transforms as

$$\tilde{g}_{\mu\nu} = \frac{\partial x^\alpha}{\partial \tilde{x}^\mu} \frac{\partial x^\beta}{\partial \tilde{x}^\nu} g_{\alpha\beta}. \quad (4.5)$$

If we perform a purely spatial transformation, $\tilde{x}^0 = x^0$, $\tilde{x}^i = f(x^i)$, we have immediately that $\tilde{g}_{00} = g_{00}$, as requested for a spatial scalar.

If we write the perturbed metric as $g_{\mu\nu} = g_{\mu\nu}^{(0)} + \delta g_{\mu\nu}$, the condition that $g_{\alpha\gamma} g^{\gamma\beta} = \delta_\alpha^\beta$ imposes the following relation at first order:

$$\delta g^{\mu\nu} = -\delta g_{\alpha\beta} g^{(0)\alpha\mu} g^{(0)\beta\nu}. \quad (4.6)$$

That is, the inverse of the perturbed metric is minus the perturbed metric with indices raised by the unperturbed metric.

A decomposition analogous to $g_{\mu\nu}$ can be done for any rank-two tensor as e.g., the energy-momentum tensor. From Helmholtz's theorem one can decompose the vector w_i into a longitudinal and a transverse component

$$w_i = w_i^\parallel + w_i^\perp, \quad (4.7)$$

where by construction

$$\nabla \cdot w_i^\perp = \nabla \times w_i^\parallel = 0. \quad (4.8)$$

The longitudinal component, $w_i^\parallel$, being curl-free, is the gradient of a scalar quantity w_s , i.e. $w_i^\parallel = \nabla w_s$. When we derive the Einstein equations for the $(0i)$ components, we will have therefore longitudinal and transverse terms, both in G_{0i} and in T_{0i} .

Taking the curl of the equations, we are left with only the transverse equations. On the other hand, taking the divergence, we are left with the longitudinal ones. Therefore, the two components completely *decouple* from each other and evolve independently, and therefore can be treated separately. Since the density perturbation δ is a scalar quantity, only the longitudinal terms, which can be derived from a scalar quantity, couple to the density perturbations.

A similar argument holds for the traceless spatial part h_{ij} . This tensor can be written in general as a sum of three traceless terms:

$$h_{ij} = h_{ij}^{\parallel} + h_{ij}^{\perp} + h_{ij}^T, \quad (4.9)$$

where the divergences $\partial^i h_{ij}^{\parallel}$, $\partial^i h_{ij}^{\perp}$ (which are vectors) are longitudinal and transverse, respectively, and h_{ij}^T is transverse, that is

$$\epsilon_{ijk} \partial_i \partial_k h_{ij}^{\parallel} = 0, \quad \partial_i \partial_j h_{ij}^{\perp} = 0, \quad \partial_i h_{ij}^T = 0. \quad (4.10)$$

Here the Levi-Civita tensor ϵ_{ijk} is +1 for even index permutations (123, 312, 231), -1 for odd permutations, and 0 for repeated indices. This means that the curl of $\partial_i h_{ij}^{\parallel}$ as well as divergences of $\partial_j h_{ij}^{\perp}$ and h_{ij}^T vanish. Now, since $\partial_i h_{ij}^{\parallel}$ is curl-free, it can be written in terms of a scalar function B , and it is easy to verify that $\epsilon_{ijk} \partial_i \partial_k h_{ij}^{\parallel} = 0$ if

$$h_{ij}^{\parallel} = \left(\partial_i \partial_j - \frac{1}{3} \delta_{ij} \nabla^2 \right) B \equiv D_{ij} B, \quad (4.11)$$

where the traceless operator D_{ij} is defined implicitly. On the contrary, the perturbations $h_{ij}^{\perp}$, h_{ij}^T cannot be derived from a scalar function. The first one is a vector giving rise to rotational velocity perturbations, whereas the second one is a tensor giving rise to gravitational waves. Both decouple completely from the scalar terms and can be treated separately. The terms which are intrinsically vectorial couple to pure rotational modes, while tensorial terms represent gravitational waves, coupled to matter only for anisotropic perturbations. Furthermore, it can be shown that if initially the rotational, or vorticity, modes are zero, they remain zero throughout (unless there are entropy gradients). If they are present initially, they decrease as a^{-1} . Since they are not of much interest in current dark energy research we will not further deal with vector or tensor perturbations.

Therefore, we need to take into account only the part of w_i and h_{ij} derived from scalars. This may be done by introducing two new scalar functions, E and B , that produce the vector $E_{,i}$ and the tensor $D_{ij} B$, in analogy to the electromagnetic forces. Then the perturbed metric is given by

$$\delta g_{\mu\nu} = a^2 \begin{pmatrix} -2\Psi & E_{,i} \\ E_{,i} & 2\Phi\delta_{ij} + D_{ij} B \end{pmatrix}. \quad (4.12)$$

Out of the four scalar functions Ψ , Φ , E , B , one can construct gauge-invariant quantities, that is, combinations that remain invariant at first-order under a general coordinate infinitesimal transformation, $\tilde{x}^\mu = x^\mu + \xi^\mu$. As we have already mentioned, the situation can be much simplified if one works in a specific gauge. This can be done by imposing up to four conditions on the metric, which corresponds to the four gauge coordinate transformations. Here we choose them to be $w_i = 0$ (from which $E = 0$) and $B = 0$. This finally leaves the perturbed metric in the *Newtonian or longitudinal or shear-free* gauge:

$$ds^2 = a^2(\eta) \left[-(1 + 2\Psi)d\eta^2 + (1 + 2\Phi)\delta_{ij}dx^i dx^j \right]. \quad (4.13)$$

Beware of the signs: many authors choose opposite metric signature and/or opposite signs for Φ and Ψ , which can lead to a great deal of confusion. We follow the choice of Dodelson's textbook [74].

In order to derive the first-order Einstein equations, we decompose the Einstein tensor G^μ_ν and the energy-momentum tensor T^μ_ν into background and perturbed parts: $G^\mu_\nu = G^\mu_{\nu(0)} + \delta G^\mu_\nu$ and $T^\mu_\nu = T^\mu_{\nu(0)} + \delta T^\mu_\nu$. The background cosmological evolution is obtained by solving the zero-th order Einstein equations, $G^\mu_{\nu(0)} = 8\pi G T^\mu_{\nu(0)}$. The first-order Einstein equations are given by

$$\delta G^\mu_\nu = 8\pi G \delta T^\mu_\nu. \quad (4.14)$$

The l.h.s. of Eq. (4.14) can be computed by the following procedure. We first need to calculate the perturbed Christoffel symbols $\delta\Gamma^\mu_{\nu\lambda}$ by using the formula:

$$\delta\Gamma^\mu_{\nu\lambda} = \frac{1}{2}\delta g^{\mu\alpha}(g_{\alpha\nu,\lambda} + g_{\alpha\lambda,\nu} - g_{\nu\lambda,\alpha}) + \frac{1}{2}g^{\mu\alpha}(\delta g_{\alpha\nu,\lambda} + \delta g_{\alpha\lambda,\nu} - \delta g_{\nu\lambda,\alpha}). \quad (4.15)$$

For the metric (4.13), the non-vanishing components of perturbed Christoffel symbols are

$$\delta\Gamma^0_{ij} = \delta_{ij} [2\mathcal{H}(\Phi - \Psi) + \Phi'], \quad (4.16)$$

$$\delta\Gamma^0_{00} = \Psi', \quad (4.17)$$

$$\delta\Gamma^0_{0i} = \delta\Gamma^i_{00} = \Psi_{,i}, \quad (4.18)$$

$$\delta\Gamma^i_{j0} = \delta^i_j \Phi'. \quad (4.19)$$

Note that in this chapter a prime represents the derivative with respect to the conformal time η apart from the use for dummy integration variables such as $\mathbf{k}'$ and r' .

The next step is to derive the perturbations in the Ricci tensor and in the Ricci scalar:

$$\delta R_{\mu\nu} = \delta \Gamma_{\mu\nu,\alpha}^{\alpha} - \delta \Gamma_{\mu\alpha,\nu}^{\alpha} + \delta \Gamma_{\mu\nu}^{\alpha} \Gamma_{\alpha\beta}^{\beta} + \Gamma_{\mu\nu}^{\alpha} \delta \Gamma_{\alpha\beta}^{\beta} - \delta \Gamma_{\mu\beta}^{\alpha} \Gamma_{\alpha\nu}^{\beta} - \Gamma_{\mu\beta}^{\alpha} \delta \Gamma_{\alpha\nu}^{\beta}, \quad (4.20)$$

$$\delta R = \delta g^{\mu\alpha} R_{\alpha\mu} + g^{\mu\alpha} \delta R_{\alpha\mu}. \quad (4.21)$$

Finally the perturbed Einstein tensors are derived by

$$\delta G_{\mu\nu} = \delta R_{\mu\nu} - \frac{1}{2} \delta g_{\mu\nu} R - \frac{1}{2} g_{\mu\nu} \delta R, \quad (4.22)$$

$$\delta G_{\nu}^{\mu} = \delta g^{\mu\alpha} G_{\alpha\nu} + g^{\mu\alpha} \delta G_{\alpha\nu}. \quad (4.23)$$

For the metric (4.13) we obtain

$$\delta G_0^0 = 2a^{-2} [3\mathcal{H}(\mathcal{H}\Psi - \Phi') + \nabla^2 \Phi], \quad (4.24)$$

$$\delta G_i^0 = 2a^{-2} (\Phi' - \mathcal{H}\Psi)_{|i}, \quad (4.25)$$

$$\begin{aligned} \delta G_j^i = 2a^{-2} [(\mathcal{H}^2 + 2\mathcal{H}')\Psi + \mathcal{H}\Psi' - \Phi'' - 2\mathcal{H}\Phi'] \delta_j^i \\ + a^{-2} [\nabla^2(\Psi + \Phi)\delta_j^i - (\Psi + \Phi)_{|j}^i], \end{aligned} \quad (4.26)$$

where the subscript “|” represents a covariant derivative with the spatial 3-metric and $\nabla^2 f \equiv f_{;i}^{;i}$. If the matter source is specified, the perturbed energy-momentum tensor δT_{ν}^{μ} is determined accordingly. We then obtain the linear perturbation equations from Eq. (4.14). We also recall that the energy-momentum tensor satisfies the continuity equation $T_{\nu;\mu}^{\mu} = 0$. The first-order part of this equation,

$$\delta T_{\nu;\mu}^{\mu} = 0, \quad (4.27)$$

also gives a number of useful equations, as we will see later.

In order to evaluate the perturbed energy-momentum tensor, we need to perturb the four-velocity $u^{\mu} \equiv \frac{dx^{\mu}}{ds}$. Neglecting the perturbations higher than the first order, we obtain

$$\begin{aligned} u^{\mu} &= \left[\frac{1}{a} (1 - \Psi), \frac{v^i}{a} \right], \\ u_{\mu} &= g_{\mu\nu} u^{\nu} = [-a(1 + \Psi), av_i], \\ u_{\mu} u^{\mu} &= -1, \end{aligned} \quad (4.28)$$

where $v^i = \frac{dx^i}{d\eta} = a \frac{dx^i}{dt}$ is the matter peculiar velocity with respect to the general expansion.

A note about the definition of the peculiar velocity is in order here. From Eq. (2.58) the comoving distance at redshift z is given by (after setting $c = 1$

and $a_0 = 1$)

$$d_c(z) = \int_0^z \frac{d\tilde{z}}{H(\tilde{z})}. \quad (4.29)$$

However, due to the peculiar velocity of the source, the redshift is the sum of cosmological expansion and peculiar velocity Doppler effect,

$$z = z_c + z_p = z_c + \frac{v_p}{a} \quad (z_p \ll z_c), \quad (4.30)$$

where $v_p = az_p$ is the projection of the peculiar velocity vector along the line of sight. Then we have

$$d_c(z) = \int_0^{z_c+z_p} \frac{d\tilde{z}}{H(\tilde{z})} \simeq \int_0^{z_c} \frac{d\tilde{z}}{H(\tilde{z})} + \frac{z_p}{H(z_c)} = \int_0^{z_c} \frac{d\tilde{z}}{H(\tilde{z})} + \frac{(1+z_c)v_p}{H(z_c)}. \quad (4.31)$$

The extra apparent distance induced by the peculiar velocity is therefore $(1+z)H(z)^{-1}v_p$. This can be always neglected, except when one is interested in the properties of the redshift space/real space conversion, see Section 14.3.

4.3 Single-fluid model

Let us consider a single-fluid model with an energy-momentum tensor $T_{\mu\nu}$. For a general fluid the energy momentum tensor is given by

$$T_{\mu\nu} = (\rho + P)u_\mu u_\nu + Pg_{\mu\nu} + [q_\mu u_\nu + q_\nu u_\mu + \pi_{\mu\nu}], \quad (4.32)$$

where, beside the familiar symbols ρ , P , u_μ for the energy density, the pressure and the four-velocity vector, we meet the heat flux vector q_μ and the viscous shear tensor $\pi_{\mu\nu}$. The terms inside square brackets in Eq. (4.32) are important only for fluids whose internal energy is a sizable fraction of the total energy density. We have $q_\mu = 0$ and $\pi_{\mu\nu} = 0$ for perfect fluids. In the following we limit ourselves to perfect fluids. We also assume that the *perturbed* fluid remains a perfect fluid. This implies that $\Sigma_j^i \equiv \delta T_j^i = 0$ ($i \neq j$), a condition that will be used below.

The notation for the perturbed quantities is

$$\delta \equiv \frac{\delta\rho}{\rho}, \quad \theta \equiv \nabla_i v^i, \quad (4.33)$$

where $\delta\rho/\rho \equiv (\rho(x) - \bar{\rho})/\bar{\rho}$ is the density contrast ($\rho(x)$ is the density field and $\bar{\rho}$ is the spatial average) and θ is the velocity divergence. In general there are several pairs δ_i , θ_i , one for each perfect fluid composing the Universe. All of the perturbed quantities are functions of space x and time t .

A simple observation can avoid much confusion later on: we will often speak loosely of $\delta(x)$ as if it were a deterministic quantity and we will say that δ grows

or decays. However $\delta(x)$ is in reality a random field which by definition has a zero mean value $\langle \delta \rangle = 0$. What we mean is that in the linear regime the value of $\delta(x)$ at any point x grows or decays so that $\delta(x, t) = D(t)\delta(x, 0)$, where $D(t)$ is the growth (or decay) function. In the linear regime the spatial part is always factored out and its properties are assigned by initial conditions. We will always assume Gaussian initial conditions as predicted in standard inflationary models.

From Eq. (4.32) the perturbed energy-momentum tensor for a perfect fluid with the equation of state $w = P/\rho$ can be written as

$$\delta T_\nu^\mu = \rho[\delta(1 + c_s^2)u_\nu u^\mu + (1 + w)(\delta u_\nu u^\mu + u_\nu \delta u^\mu) + c_s^2 \delta \delta_\nu^\mu], \quad (4.34)$$

where δ_ν^μ should not be confused with the density contrast δ . Here we have introduced the sound velocity, $c_s^2 \equiv \delta P / \delta \rho$. If P , even when perturbed, depends on ρ alone (which is the case called *barotropic fluid*) then

$$c_s^2 \equiv \frac{\delta P}{\delta \rho} = \frac{dP}{d\rho} = \frac{\dot{P}}{\dot{\rho}}. \quad (4.35)$$

The last passage is valid only in the FLRW metric where at background level everything depends on time alone (c_s is calculated at zero-th order since it will always appear as a factor of first-order variables). Since c_s , just as w , depends at first-order only on background quantities, in this case the perturbation equations do not introduce any new free function. In general, however, the pressure P can depend on internal degrees of freedom of the fluid, say, entropy s . Then one has

$$c_s^2 = \frac{\delta P(\rho, s)}{\delta \rho} = \frac{\partial P}{\partial \rho} + \frac{\partial P}{\partial s} \frac{\partial s}{\partial \rho} = c_{s(a)}^2 + c_{s(na)}^2, \quad (4.36)$$

where $c_{s(a)} \equiv \sqrt{\bar{P}/\dot{\rho}}$ is called the *adiabatic* sound speed and $c_{s(na)}$ is, guess what, the non-adiabatic sound speed. The non-adiabatic sound speed in general will depend on microphysical properties of the fluid and appears as a new free function only at the level of perturbations. The gravitational equations at first-order are then completely specified only if we give for each fluid the equation of state $w(a)$ and the total sound speed $c_s(a)$ or, equivalently, if we assign to the fluid a function $P(\rho, s)$ which determines both.

The components of the energy-momentum tensor are

$$\delta T_0^0 = -\delta \rho, \quad (4.37)$$

$$\delta T_i^0 = -\delta T_0^i = (1 + w)\rho v^i, \quad (4.38)$$

$$\delta T_1^1 = \delta T_2^2 = \delta T_3^3 = c_s^2 \delta \rho. \quad (4.39)$$

Then the perturbed Einstein equations (4.14) lead to

$$3\mathcal{H}(\mathcal{H}\Psi - \Phi') + \nabla^2\Phi = -4\pi G a^2 \delta\rho, \quad (4.40)$$

$$\nabla^2(\Phi' - \mathcal{H}\Psi) = 4\pi G a^2 (1 + w) \rho \theta, \quad (4.41)$$

$$\Psi = -\Phi, \quad (4.42)$$

$$\Phi'' + 2\mathcal{H}\Phi' - \mathcal{H}\Psi' - (\mathcal{H}^2 + 2\mathcal{H}')\Psi = -4\pi G a^2 c_s^2 \delta\rho. \quad (4.43)$$

Note that Eqs. (4.40)–(4.43) come from the (00), (0i), (ij), and (ii) components. Equation (4.42) follows from the property $\delta T_j^i = 0$.

One can also derive some useful equations by using the continuity equation (4.27). Recall that the operation of covariant divergence of a tensor T_ν^μ is

$$T_{\nu;\mu}^\mu = T_{\nu,\mu}^\mu - \Gamma_{\nu\beta}^\alpha T_\alpha^\beta + \Gamma_{\beta\alpha}^\alpha T_\nu^\beta. \quad (4.44)$$

Then the $\nu = 0$ component of Eq. (4.27), i.e. $\delta T_{0;\mu}^\mu = 0$, reads

$$\delta T_{0,\mu}^\mu - \delta \Gamma_{0\beta}^\alpha T_\alpha^\beta - \Gamma_{0\beta}^\alpha \delta T_\alpha^\beta + \delta \Gamma_{0\alpha}^\alpha T_0^0 + \Gamma_{\beta\alpha}^\alpha \delta T_0^\beta = 0, \quad (4.45)$$

which reduces to

$$(\delta\rho)' + 3\mathcal{H}(\delta\rho + \delta P) = -(\rho + P)(\theta + 3\Phi'), \quad (4.46)$$

where we have employed Eqs. (4.16)–(4.19). Using the unperturbed conservation equation $\rho' + 3\mathcal{H}(\rho + P) = 0$ together with the relations $w = P/\rho$ and $c_s^2 = \delta P/\delta\rho$, we find that Eq. (4.46) can be expressed as

$$\delta' + 3\mathcal{H}(c_s^2 - w)\delta = -(1 + w)(\theta + 3\Phi'), \quad (4.47)$$

which is called the (perturbed) *continuity equation*. For non-relativistic matter with $w = 0$ and $c_s^2 = 0$, this equation reduces to

$$\delta' = -\theta - 3\Phi' \quad (\text{for non-relativistic matter}). \quad (4.48)$$

This equation tells us that the density at position x increases if there is a velocity divergence in the same place, that is, if there is more matter coming in than going out. The Φ' term, absent in Newtonian dynamics, is negligible at small scales and, of course, for a slowly varying gravitational potential.

The equation $\delta T_{\nu;\mu}^\mu = 0$ for $\nu = i$ leads to

$$\delta q' + 3\mathcal{H}\delta q = -a\delta P - (\rho + P)a\Psi, \quad (4.49)$$

where $\delta q \equiv a(\rho + P)v$ and v is a velocity potential related to v^i via $v^i = \nabla^i v$. Writing Eq. (4.49) in terms of v^i and taking the divergence ∇_i , we obtain

$$\theta' + \left[\mathcal{H}(1 - 3w) + \frac{w'}{1 + w} \right] \theta = -\nabla^2 \left(\frac{c_s^2}{1 + w} \delta + \Psi \right). \quad (4.50)$$

For non-relativistic matter, this reduces to

$$\theta' + \mathcal{H}\theta = -\nabla^2\Psi - \nabla^2(c_s^2\delta) \quad (\text{for non-relativistic matter}), \quad (4.51)$$

where now we have included the $\nabla^2(c_s^2\delta)$ term. This is called the *Euler equation* in the Newtonian context. It says that the (peculiar) acceleration depends on the sum of the potential and pressure gradients.

We go now to the *Fourier space*. This means that all perturbation quantities will be Fourier expanded (we do not need to bother with pre-factors here):

$$\Phi = \int e^{i\mathbf{k}\cdot\mathbf{r}} \Phi_k d^3k, \quad \Psi = \int e^{i\mathbf{k}\cdot\mathbf{r}} \Psi_k d^3k, \quad (4.52)$$

$$\delta = \int e^{i\mathbf{k}\cdot\mathbf{r}} \delta_k d^3k, \quad \theta = \int e^{i\mathbf{k}\cdot\mathbf{r}} \theta_k d^3k. \quad (4.53)$$

The subscript k represents a Fourier mode for each wavenumber k . Note that k is a comoving quantity that remains fixed. In the following we drop the subscript k as long as no confusion arises by doing so. In Fourier space we assume that the perturbation variables ($\delta, \theta, \Psi, \Phi$ etc.) are the sum of plane waves $\delta_k e^{i\mathbf{k}\cdot\mathbf{r}}$. Since the equations are linear, each plane wave obeys the same equations with a different comoving wavenumber k . Throughout the linear evolution, the physical scale λ_p of the perturbation expands with the cosmic expansion as $\lambda_p = (2\pi/k)a$. Of course, if the perturbation enters a non-linear regime, then this treatment breaks down and the perturbation decouples from the Hubble expansion and starts collapsing. When we calculate the perturbation equations with an algebraic manipulator it is very convenient to introduce from the beginning all perturbation variables as Fourier modes, e.g., $\delta(x, y, z, t) = \delta_k(t) e^{i\mathbf{k}\cdot\mathbf{r}}$. Since we are always interested in the direction-averaged equations (i.e. the equations that depend only on the modulus k), we could simply put $\mathbf{k} \cdot \mathbf{r} = k(x + y + z)/\sqrt{3}$.

In practice, each perturbation quantity ϕ and its derivatives can be substituted as follows

$$\phi(x, \eta) \rightarrow e^{i\mathbf{k}\cdot\mathbf{r}} \phi(\eta), \quad (4.54)$$

$$\nabla\phi(x, \eta) \rightarrow i e^{i\mathbf{k}\cdot\mathbf{r}} \mathbf{k} \phi(\eta), \quad (4.55)$$

$$\nabla^2\phi(x, \eta) \equiv \nabla_i \nabla^i \phi(x, \eta) \rightarrow -e^{i\mathbf{k}\cdot\mathbf{r}} k^2 \phi(\eta). \quad (4.56)$$

Notice that when there are two repeated spatial indices we sum over them without the help of the metric coefficients g_{ij} (more exactly, we use the induced 3-dimensional spatial metric which for spatially flat spaces is just the Euclidean metric). Furthermore, the Fourier modes $e^{i\mathbf{k}\cdot\mathbf{r}}$ can be simply dropped out, since the equations are linear and therefore decoupled between different modes.

From Eqs. (4.40)–(4.43), (4.47), and (4.50) we obtain the following equations for each Fourier mode:

$$k^2\Phi + 3\mathcal{H}(\Phi' - \mathcal{H}\Psi) = 4\pi G a^2 \rho \delta, \quad (4.57)$$

$$k^2(\Phi' - \mathcal{H}\Psi) = -4\pi G a^2(1 + w)\rho\theta, \quad (4.58)$$

$$\Psi = -\Phi, \quad (4.59)$$

$$\Phi'' + 2\mathcal{H}\Phi' - \mathcal{H}\Psi' - (\mathcal{H}^2 + 2\mathcal{H}')\Psi = -4\pi G a^2 c_s^2 \rho \delta, \quad (4.60)$$

$$\delta' + 3\mathcal{H}(c_s^2 - w)\delta = -(1 + w)(\theta + 3\Phi'), \quad (4.61)$$

$$\theta' + \left[\mathcal{H}(1 - 3w) + \frac{w'}{1 + w} \right] \theta = k^2 \left(\frac{c_s^2}{1 + w} \delta + \Psi \right), \quad (4.62)$$

where now

$$\theta = i\mathbf{k} \cdot \mathbf{v}. \quad (4.63)$$

Although the six equations above are not independent they are all useful. Let us remark again that we are considering here the fluctuations of a Universe composed of a *single fluid*. In the next sections we will solve the general set of equations (4.57)–(4.62) in several different regimes: during the epochs of radiation and matter eras, and both at large and small scales.

Finally, one can combine Eqs. (4.57) and (4.58) to get the relativistic Poisson equation

$$k^2\Phi = 4\pi G a^2 \rho [\delta + 3\mathcal{H}(w + 1)\theta/k^2] = 4\pi G a^2 \rho \delta^*, \quad (4.64)$$

where we define the *total-matter* variable:

$$\delta^* \equiv \delta + 3\mathcal{H}(w + 1)\theta/k^2. \quad (4.65)$$

Notice that in our conventions an overdensity $\delta^* > 0$ generates opposite gravitational potentials, $\Phi > 0$ and $\Psi < 0$, on small scales.

Combining Eqs. (4.57), (4.59), and (4.60), we can get an equation for Φ alone:

$$\Phi'' + 3\mathcal{H}(1 + c_s^2)\Phi' + (c_s^2 k^2 + 3\mathcal{H}^2 c_s^2 + 2\mathcal{H}' + \mathcal{H}^2)\Phi = 0. \quad (4.66)$$

Similarly, using the relativistic Poisson equation (4.64), this becomes an equation for δ^* :

$$(\delta^*)'' + \mathcal{H}(1 + 3c_s^2 - 6w)(\delta^*)' - \left[\frac{3}{2}\mathcal{H}^2(1 - 6c_s^2 - 3w^2 + 8w) - c_s^2 k^2 \right] \delta^* = 0, \quad (4.67)$$

where c_s , w are arbitrary functions of time and where we have employed the useful relation

$$\mathcal{H}' = -\frac{1}{2}(1 + 3w)\mathcal{H}^2. \quad (4.68)$$

4.4 Scales larger than the horizon

Now that we have derived the perturbation equations, we can begin wondering how to solve them. As a first example we work out the simplest case, the large-scale limit $k \ll \mathcal{H} = aH$. This corresponds to the scale on which the physical wavelength $\lambda_p = (2\pi/k)a$ of perturbations is much larger than the Hubble radius H^{-1} , i.e. *super-horizon scales* (although notice that the horizon corresponds approximately to $1/H$ only for some particular case). If the pressure depends only on the energy density and the equation of state w is a constant then we have $c_s^2 = w$, which is valid both for matter and radiation. In this case Eq. (4.66), using Eq. (4.68), reduces to

$$\Phi'' + 3\mathcal{H}(1 + c_s^2)\Phi' = 0. \quad (4.69)$$

Then $\Phi' = 0$ is a solution. Equation (4.57) becomes

$$3\mathcal{H}^2\Phi = 4\pi G a^2 \rho \delta. \quad (4.70)$$

Using the Friedmann equation, $3\mathcal{H}^2 = 8\pi G \rho a^2$, it follows that

$$\delta = 2\Phi. \quad (4.71)$$

Hence $\Phi = \text{constant}$ at large scales implies that $\delta = \text{constant}$. One easily finds that the result (4.71) is consistent with the other Einstein equations. Equation (4.69) is second-order, so we must have two solutions. It appears immediately that $\Phi = \text{constant}$ is a growing mode or a dominating solution (at least for $c_s^2 > -1$). Thus we have shown that the gravitational potential remains constant for scales outside the Hubble radius whenever $c_s^2 = w$ for the total fluid. During the transition from radiation to matter eras this condition is violated and as we will see in Section 4.10 the gravitational potential changes.

4.5 Scales smaller than the Hubble radius

Next, we work out the opposite case, $k \gg \mathcal{H}$, i.e. scales deep inside the Hubble radius (*sub-horizon scales*). The fluctuation of a pressureless fluid can grow indefinitely because there is no counteracting force. In general, however, the pressure of the fluid resists gravity and stops the collapse. It is then instructive to derive the equations for a fluid which is pressureless ($w = 0$) in the absence of perturbations,

but has a small sound speed:

$$c_s^2 = \frac{\delta P}{\delta \rho} \ll 1. \quad (4.72)$$

Then Eq. (4.58) tells us that $\Phi' - \mathcal{H}\Psi \simeq 0$, so that Eq. (4.57) corresponds to the Fourier transformed Poisson equation

$$k^2 \Phi = 4\pi G a^2 \rho \delta = \frac{3}{2} \mathcal{H}^2 \delta. \quad (4.73)$$

Taking the derivative of Eq. (4.73) and substituting it into Eq. (4.61), we obtain

$$\delta' = -\theta - \frac{9}{2} \frac{\mathcal{H}^2}{k^2} \delta \left(2 \frac{\mathcal{H}'}{\mathcal{H}} + \frac{\delta'}{\delta} \right) \simeq -\theta. \quad (4.74)$$

Hence this equation reduces to the energy conservation equation in the Newtonian limit.

Then the perturbation equations in the sub-horizon limit become

$$\delta' = -\theta, \quad (4.75)$$

$$\theta' = -\mathcal{H}\theta + c_s^2 k^2 \delta - k^2 \Phi, \quad (4.76)$$

and Eq. (4.73). Differentiating Eq. (4.75) with respect to η and using Eq. (4.76), it follows that

$$\delta'' + \mathcal{H}\delta' + \left(c_s^2 k^2 - \frac{3}{2} \mathcal{H}^2 \right) \delta = 0. \quad (4.77)$$

In the Minkowski limit, $\mathcal{H} \rightarrow 0$, this equation reduces to the classical fluid wave equation $\delta'' + c_s^2 k^2 \delta = 0$, where c_s is indeed the sound velocity. Equation (4.77) shows at once that the perturbation does not grow if

$$c_s^2 k^2 - \frac{3}{2} \mathcal{H}^2 > 0, \quad (4.78)$$

i.e. if the physical wavelength $\lambda_p = (2\pi/k)a$ is smaller than the Jeans length,

$$\lambda_J = c_s \sqrt{\frac{\pi}{G\rho}}. \quad (4.79)$$

For scales smaller than λ_J the perturbations undergo damped oscillations. For the CDM particles the velocity dispersion is always negligible, at least in the regime of validity of our linear treatment. For the photons we have $c_s = c/\sqrt{3}$, so that

$$\lambda_J \approx H^{-1}. \quad (4.80)$$

Hence the growth of perturbations is prevented on all scales smaller than the Hubble radius. For the baryons, the sound velocity is comparable to the photon

velocity before the decoupling epoch, so that baryon perturbations are damped out (more precisely they drop rapidly to a comoving scale of less than 1 Mpc just after decoupling). Then the baryons are free to fall inside the dark matter potential wells, and their perturbation spectrum catches the dark matter one (more on this in the next section).

When $c_s k \ll \mathcal{H}$, the perturbations grow freely because gravity overcomes the pressure: this is the all-important regime of *gravitational instability*. The sub-horizon equation for a single pressureless fluid becomes

$$\delta'' + \mathcal{H}\delta' - \frac{3}{2}\mathcal{H}^2\delta = 0. \quad (4.81)$$

It is often useful to employ the number of e-foldings $N = \ln a$. Then Eq. (4.81) can be written as

$$\frac{d^2\delta}{dN^2} + \left(\frac{1}{\mathcal{H}} \frac{d\mathcal{H}}{dN} + 1 \right) \frac{d\delta}{dN} - \frac{3}{2}\delta = 0. \quad (4.82)$$

We can rewrite Eq. (4.68) as

$$\frac{1}{\mathcal{H}} \frac{d\mathcal{H}}{dN} = -\frac{1}{2} - \frac{3}{2}w. \quad (4.83)$$

For a pressureless fluid ($w = 0$), Eq. (4.82) then reduces to

$$\frac{d^2\delta}{dN^2} + \frac{1}{2} \frac{d\delta}{dN} - \frac{3}{2}\delta = 0. \quad (4.84)$$

The direct substitution $\delta = Ae^{\lambda N}$ gives the solution $\lambda = 1, -3/2$. Then the evolution of growing and decaying modes during the matter era is given by

$$\delta_+ = Aa, \quad \delta_- = Ba^{-3/2}. \quad (4.85)$$

In terms of the cosmic time t , the growing solution evolves as $\delta_+ \propto t^{2/3}$. The pre-factor is of course fixed by the initial conditions, ultimately established during inflation. The decaying solutions (or in general the slower one) become soon negligible with respect to the growing ones and we will systematically neglect them throughout this book.

Inserting δ_+ into the Poisson equation (4.73), we see that $\Phi \propto a^2 H^2 \delta_+ \propto a^2 a^{-3} a^1 \propto \text{constant}$. Hence the gravitational potential remains constant during the pure matter-dominated epoch.

4.6 Two-fluid solutions

We generalize the single-fluid case to the more realistic case in which both matter ($w_m = c_s^2 = 0$) and radiation ($w_r = c_s^2 = 1/3$) are present. We introduce the matter

perturbation variables δ_m, θ_m and the radiation perturbation variables δ_r, θ_r (here radiation means all the components which are massless or relativistic, in particular the neutrinos). Since we are considering dark matter as a dominant matter component, there is no explicit interaction term between matter and radiation. The baryonic fraction is also effectively decoupled after $z \approx 1000$, while before this epoch it can be considered as a part of a relativistic photon–baryon plasma. In Fourier space we then have a system of gravitationally coupled equations for the perturbations on sub-horizon scales:

$$\delta'_m = -(\theta_m + 3\Phi'), \quad (4.86)$$

$$\theta'_m = -\mathcal{H}\theta_m - k^2\Phi, \quad (4.87)$$

$$\delta'_r = -\frac{4}{3}(\theta_r + 3\Phi'), \quad (4.88)$$

$$\theta'_r = k^2 \left(\frac{3}{4} c_s^2 \delta_r - \Phi \right), \quad (4.89)$$

$$k^2(\Phi' + \mathcal{H}\Phi) = -4\pi G(1 + w_{\text{eff}})a^2 \rho_t \theta_t, \quad (4.90)$$

$$k^2\Phi + 3\mathcal{H}(\Phi' + \mathcal{H}\Phi) = 4\pi G a^2 \rho_t \delta_t. \quad (4.91)$$

The subscript t represents total perturbation variables, i.e.

$$\rho_t = \rho_m + \rho_r, \quad (4.92)$$

$$w_{\text{eff}} = \Omega_r w_r + \Omega_m w_m = \frac{\rho_r/3}{\rho_m + \rho_r}, \quad (4.93)$$

$$\theta_t = \frac{(1 + w_m)\Omega_m \theta_m + (1 + w_r)\Omega_r \theta_r}{1 + w_{\text{eff}}}, \quad (4.94)$$

$$\delta_t = \Omega_m \delta_m + \Omega_r \delta_r. \quad (4.95)$$

Here the total effective equation of state $w_{\text{eff}} = P_t/\rho_t$ is given by

$$w_{\text{eff}} = -1 - \frac{2}{3} \frac{\dot{H}}{H^2}, \quad (4.96)$$

which follows from Eqs. (2.17)–(2.18) and $K = 0$.

We remind the reader that Ω_m and Ω_r are functions of time. These should be distinguished from their present values $\Omega_m^{(0)}$ and $\Omega_r^{(0)}$. In the sub-horizon limit, Eq. (4.91) gives

$$k^2\Phi \simeq 4\pi G a^2 (\rho_m \delta_m + \rho_r \delta_r) = \frac{3}{2} \mathcal{H}^2 (\Omega_m \delta_m + \Omega_r \rho_r). \quad (4.97)$$

Following the derivation similar to Eq. (4.77), we obtain the following equations for sub-horizon perturbations

$$\delta_m'' + \mathcal{H}\delta_m' - \frac{3}{2}\mathcal{H}^2(\Omega_m\delta_m + \Omega_r\delta_r) = 0, \quad (4.98)$$

$$\delta_r'' + \frac{k^2}{3}\delta_r = 0. \quad (4.99)$$

During the radiation-dominated epoch we have $\Omega_m \simeq 0$ and $\Omega_r \simeq 1$. Moreover the second equation shows that the radiation density contrast oscillates rapidly around zero (since we are considering sub-horizon modes, $k \gg \mathcal{H}$). The same is true for the coupled baryon–photon plasma. Therefore, we can average over the radiation oscillations and put $\langle \delta_r \rangle \simeq 0$ in the first equation. It then follows that $\Omega_m\delta_m + \Omega_r\delta_r \simeq 0$ and

$$\delta_m'' + \mathcal{H}\delta_m' \simeq 0. \quad (4.100)$$

The solution of this equation is given by $\delta_m = C_1 + C_2 \int d\eta/a$. During the radiation era the integral $\int d\eta/a$ gives only a logarithmic correction, so the matter perturbations evolve only mildly.

During the matter era we have $|\Omega_m\delta_m| \gg |\Omega_r\delta_r|$ in Eq. (4.98), so that the evolution of matter perturbations is described by $\delta_m \propto a$ as we have explained in the previous section.

If we consider cold dark matter (perturbation δ_c) and baryonic matter (perturbation δ_b) instead of matter and radiation, Eq. (4.98) can be generalized as

$$\delta_c'' + \mathcal{H}\delta_c' - \frac{3}{2}\mathcal{H}^2(\Omega_c\delta_c + \Omega_b\delta_b) = 0, \quad (4.101)$$

$$\delta_b'' + \mathcal{H}\delta_b' - \frac{3}{2}\mathcal{H}^2(\Omega_c\delta_c + \Omega_b\delta_b) = 0. \quad (4.102)$$

Since baryons correspond to a small fraction of the total matter fluid, we can assume $|\Omega_b\delta_b| \ll |\Omega_c\delta_c|$. This shows that Eq. (4.101) decouples from δ_b and reduces to the standard equation for matter perturbations.

At the same time the baryon equation is “forced” by the term $\Omega_c\delta_c$. For such coupled differential equations the asymptotic solution of δ_b will approach the forcing term δ_c . In other words, the perturbations in baryons will catch up with those in dark matter. This expresses mathematically (in the linear regime) the common expression according to which the baryons fall into the dark matter potential wells.

Analogously, if we consider the sum of pressureless matter and the cosmological constant Λ instead of matter and radiation, we get the term $\Omega_\Lambda\delta_\Lambda$ in addition to $\Omega_m\delta_m$. However ρ_Λ is constant by definition and $\delta_\Lambda = 0$, so that we have a slight

modification of Eq. (4.81):

$$\delta_m'' + \mathcal{H}\delta_m' - \frac{3}{2}\mathcal{H}^2\Omega_m\delta_m = 0. \quad (4.103)$$

This equation can be rewritten in terms of the derivative N :

$$\frac{d^2\delta_m}{dN^2} + \left(\frac{1}{\mathcal{H}} \frac{d\mathcal{H}}{dN} + 1 \right) \frac{d\delta_m}{dN} - \frac{3}{2}\Omega_m\delta_m = 0. \quad (4.104)$$

If we assume that $\Omega_m = \text{constant}$, then the solution is given by $\delta_m \sim a^{\lambda_{\pm}}$ with

$$\lambda_{\pm} = \frac{1}{4} \left(-1 \pm \sqrt{1 + 24\Omega_m} \right). \quad (4.105)$$

This case occurs when the fraction $1 - \Omega_m$ is into some form of energy density which has $w \approx 0$ but contrary to ordinary CDM it does not cluster on sub-horizon scales. The major example of this is massive neutrinos after they become non-relativistic. Supposing for a moment we could apply it also for Λ CDM, this would show that the cosmological constant slows down the perturbation growth. In the limit $\Omega_m \rightarrow 0$ we have $\lambda_+ \rightarrow 0$ from Eq. (4.105), which is qualitatively correct. However, the density parameter

$$\Omega_m = \frac{\rho_m}{\rho_m + \rho_{\Lambda}} = \frac{\rho_m^{(0)} a^{-3}}{\rho_m^{(0)} a^{-3} + \rho_{\Lambda}} \quad (4.106)$$

is obviously not a constant. A much better approximation, obtained by an empirical fit, is given by defining the *growth rate* f of matter perturbations [95]:

$$f \equiv \frac{d \ln \delta_m}{d \ln a} = \Omega_m^{\gamma}, \quad (4.107)$$

that is

$$\delta_m(a) = \delta_m(a_i) \exp \left(\int_{a_i}^a \Omega_m(\tilde{a})^{\gamma} \frac{d\tilde{a}}{\tilde{a}} \right), \quad (4.108)$$

where the growth index γ is ≈ 0.55 for the Λ CDM model (see Section 11.1 for the derivation and for a generalization to dark energy). With this behavior we realize that the term $\mathcal{H}^2\delta_m$ in the Poisson equation is no longer constant and therefore the gravitational potential Φ on sub-horizon scales is not constant. For the Λ CDM model the gravitational potential is almost constant during the matter era, but it begins to decrease after the universe enters the dark-energy-dominated epoch.

4.7 Velocity field

The mass power spectrum can be studied also by analyzing the peculiar motion of the galaxies. It is intuitive, in fact, that a more clustered distribution of matter will

induce stronger peculiar velocities. The importance of this is that the velocity field depends on the *total* mass distribution, including any unseen component. Let us start from Eq. (4.75) with $\theta = \nabla_i v^i$. In Fourier space this can be written as

$$\delta'_k = -ik_i v^i. \quad (4.109)$$

In this section we recover the subscript k for Fourier modes. This equation applies to each pressureless component such as baryons and CDM separately. However, due to the gravitational coupling, each component will be driven by a dominating density contrast, as we have seen explicitly for the baryons.

We also assume that the velocity field $\mathbf{v}$ can be represented by the *galaxy* velocity field, $\mathbf{v}_g$. This of course *assumes* that the galaxy velocities are not biased with respect to the dark matter velocity field, an expectation based on the fact that galaxies move because of the common gravitational field. Strictly speaking, this only implies that the acceleration, not the velocity, is the same, and only if all matter components share the same equation of state and sound speed. Assuming similar initial conditions, universality of the gravitational interaction, and identical equation of state and sound speed, however, the galaxies can be seen as test particles of a universal peculiar velocity field. Under these assumptions, observing the peculiar velocity field $\mathbf{v}_g$ of galaxies gives information on the *total* density contrast. So in this section we take $\mathbf{v}$ to refer to the velocity field of galaxy “test particles” and δ_k to refer to the total mass.

Consider Eq. (4.51) for $c_s = 0$:

$$v^{i'} = -\mathcal{H}v^i + ik^i \Phi_k. \quad (4.110)$$

Since we are dealing only with scalar perturbations, the velocity can be written as the gradient of a velocity potential v , i.e. $v^i = \nabla^i v \rightarrow ik^i v$. Then it is clear that v^i is parallel to k^i and we can look for solutions of Eq. (4.109) in the form $v^i = F(k, a)k^i$. This gives immediately from Eq. (4.109) the relation between the peculiar velocity field v^i and the density fluctuation δ_k in linear perturbation theory (in the Newtonian regime):

$$v^i = i\mathcal{H}f\delta_k \frac{k^i}{k^2}, \quad (4.111)$$

where f is the growth rate defined in Eq. (4.107). Substituting Eq. (4.111) into Eq. (4.109), one can easily confirm that the relation $f = \delta'_k/(\mathcal{H}\delta_k) = d \ln \delta_k / d \ln a$ follows. During the standard matter-dominated era we have already seen that $\delta_k \propto a$ and hence $f = 1$ while more in general $f = \Omega_m(a)^\gamma$.

Let us consider the present epoch (and motion on a local scale, say, less than 100 Mpc). Since $a = a_0 = 1$ and hence $\mathcal{H} = H_0$, we have

$$\mathbf{v} = i H_0 f \delta_k \frac{\mathbf{k}}{k^2}. \quad (4.112)$$

The peculiar velocity $\mathbf{v}(\mathbf{r})$ at location $\mathbf{r}$ can be obtained by Fourier anti-transformation:

$$\mathbf{v}(\mathbf{r}) = i H_0 f \frac{V}{(2\pi)^3} \int \delta_k \frac{\mathbf{k}}{k^2} e^{i\mathbf{k}\cdot\mathbf{r}} d^3k. \quad (4.113)$$

Here we implicitly assumed f to be k -independent. As we have seen this is true in Λ CDM but not necessarily in other models. The average in a volume V_R of radius R is

$$\mathbf{v}_R = i H_0 f \frac{V}{(2\pi)^3} \frac{1}{V_R} \int \delta_k \frac{\mathbf{k}}{k^2} e^{i\mathbf{k}\cdot\mathbf{r}} W(\mathbf{r}) d^3k d^3r = -i H_0 f \frac{V}{(2\pi)^3} \int \delta_k \frac{\mathbf{k}}{k^2} W(kR) d^3k. \quad (4.114)$$

Here $W(kR)$ is the Fourier transform of the *window function*, defined as

$$W(kR) \equiv \frac{1}{V_R} \int W(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}} d^3r. \quad (4.115)$$

Therefore, the average of the square of the velocity is

$$\begin{aligned} \langle v^2 \rangle_R &= H_0^2 f^2 \frac{V^2}{(2\pi)^6} \int \langle \delta_k \delta_{k'}^* \rangle \frac{\mathbf{k}}{k^2} \frac{\mathbf{k}'}{k'^2} W(kR) W(k'R) d^3k d^3k' \\ &= \frac{H_0^2 f^2}{(2\pi)^3} \int P(k) \delta_D(\mathbf{k} - \mathbf{k}') \frac{\mathbf{k}}{k^2} \frac{\mathbf{k}'}{k'^2} W(kR) W(k'R) d^3k d^3k' \\ &= \frac{H_0^2 f^2}{2\pi^2} \int P(k) W^2(kR) dk, \end{aligned} \quad (4.116)$$

where in the second line we have used the definition in Eq. (3.26) and in the last line we integrated over the solid angle 4π . The square root of $\langle v^2 \rangle_R$ is the magnitude of the peculiar flow on the scale R and is called *bulk flow*. Estimates of the bulk flow can be used to constrain or normalize the mass power spectrum (see e.g., Refs. [96, 97]). On the other hand, independent measures of $P(k)$ and $\langle v^2 \rangle_R$ can give $f^2 \simeq \Omega_m^{2\gamma}$.

4.8 The redshift distortion

The galaxy distances are mostly measured through their redshifts. However, the redshift includes the peculiar velocity of the galaxies themselves, so that there is an error in the distances we assign to galaxies. On very small scales, i.e. in the

cluster cores, the peculiar velocity of a galaxy is more or less randomly oriented, so that the error in the distance is statistical, and can be taken into account along with the experimental errors. On redshift maps, the small scale peculiar velocities cause the fingers-of-god effect: galaxies in a cluster acquire an additional random redshift that distorts the cluster distribution, making it appear elongated along the line of sight.

On large scales, however, the galaxies tend to fall toward concentrations, so that the velocity field is coupled to the density field. This correction is systematic and can be accounted for in the following way [98].

Given a peculiar velocity $\mathbf{v}$ of a source at position $\mathbf{r}$, one can define the line-of-sight component

$$u(r) \equiv \mathbf{v} \cdot \frac{\mathbf{r}}{r}, \quad (4.117)$$

where $r = |\mathbf{r}|$. The coordinate transformation from real space (r) to redshift space (s) is given by

$$s = r \left[1 + \frac{u(r) - u(0)}{r} \right]. \quad (4.118)$$

Here we express velocities in Megaparsecs through division by H_0 .

If dV_s and dV_r are the volume elements in the two coordinates with number densities $n(s)$ and $n(r)$, respectively, we have

$$n(r)dV_r = n(s)dV_s. \quad (4.119)$$

The volume element dV_s can be written in terms of the r coordinate as

$$dV_s = \left(1 + \frac{\Delta u(r)}{r} \right)^2 |J| (r^2 \sin \theta) dr d\theta d\phi = \left(1 + \frac{\Delta u(r)}{r} \right)^2 |J| dV_r, \quad (4.120)$$

where $\Delta u(r) \equiv u(r) - u(0)$. The Jacobian $|J|$ is given by the derivative of $s = r[1 + (u(r) - u(0))/r]$ with respect to r :

$$|J| = \left| \frac{\partial s}{\partial r} \right| = 1 + \frac{du}{dr}. \quad (4.121)$$

Using the average density n_0 , the density contrast in s -space is

$$\delta_s = \frac{n(s)dV_s}{n_0 dV_s} - 1 = \frac{n(r)dV_r}{n_0 dV_s} - 1 = \frac{n(r)}{n_0 (1 + \Delta u(r)/r)^2 |J|} - 1, \quad (4.122)$$

where we have used Eq. (4.120). To first order, this yields

$$\begin{aligned}\delta_s &\simeq \frac{n(r)}{n_0} \left[1 - 2 \frac{\Delta u(r)}{r} - \frac{du}{dr} \right] - 1 \\ &= \delta_r - 2 \frac{\Delta u(r)}{r} - \frac{du}{dr},\end{aligned}\tag{4.123}$$

where in the last line we have employed the fact that to first-order we can approximate $n(r)$ to be n_0 . Therefore, we see that the density contrast is different in the two spaces. As a consequence, the correlation function and the power spectrum measured in redshift space will have to be corrected to be expressed in real space. To do so, we have to take the velocity field from the linear perturbation theory.

However what we observe is the *galaxy* density contrast δ_g , which is different from the total density contrast δ_m . The two quantities can be assumed to be related by a *bias* factor b defined by

$$b \equiv \frac{\delta_g}{\delta_m},\tag{4.124}$$

so that instead of $f\delta_k$ we have $\delta_{(g)k}f/b$ in Eq. (4.112). In other words, wherever we write f , we should in fact write $\beta \equiv f/b$. If we assume b to be scale-independent, then Eq. (4.113) can be written as

$$\mathbf{v} = i H_0 \beta \int \delta_{(g)k} e^{i\mathbf{k}\cdot\mathbf{r}} \frac{\mathbf{k}}{k^2} d^3k^*,\tag{4.125}$$

where the Fourier factor $V/(2\pi)^3$ is included in the differential d^3k^* . Notice the apparent oddity: in this equation we have both the *total* δ (in β) and the *galaxy* $\delta_{(g)}$. In general however b can depend on scale and time. Notice that if b depends on space then Eq. (4.124) in real space is different from the analogous relation in Fourier space.

An immediate consequence of Eq. (4.124) written in Fourier space is that for the power spectra we have

$$P_g(k) = b^2 P_m(k).\tag{4.126}$$

One should take note however what this standard procedure is assuming. If linear gravity were the only force at work, then any bias in δ should induce a bias in $\mathbf{v}$ due to the continuity equation, while here we are implicitly assuming that *biased* galaxies possess an *unbiased* velocity field. In other words, a value $b \neq 1$ implies that galaxies violate the linear continuity equation, while dark matter still obeys it. This could only occur if some additional force acts on baryons but not on dark matter. If gravity does not violate the equivalence principle (see Section 11.2 for an example of this), one needs to invoke some non-gravitational effect, e.g.,

the hydrodynamical effect, that makes the galaxies “appear” or “disappear” and therefore breaks continuity. Examples of these are merging processes or evolutionary processes that render galaxies brighter or dimmer and therefore visible or invisible to our telescopes. We can say that the whole idea of a simple biasing scheme tries to capture glimpses of physics beyond the purely linear gravitational treatment. Ultimately, its validity can only be judged against N -body simulations and observations. In problem 4.1 you can work out an example of constant bias.

Using Eq. (4.125), the line-of-sight component (4.117) in Megaparsec units ($H_0 = 1$) is (we drop now the subscript g in δ but we add a subscript r to remind ourselves that this quantity is in real space)

$$u(r) = i\beta \int \delta_{rk} e^{ik \cdot r} \frac{\mathbf{k} \cdot \mathbf{r}}{k^2 r} d^3 k^*, \quad (4.127)$$

while its derivative is

$$\frac{du}{dr} = -\beta \int \delta_{rk} e^{ik \cdot r} \left(\frac{\mathbf{k} \cdot \mathbf{r}}{kr} \right)^2 d^3 k^*, \quad (4.128)$$

where we have used the relation

$$\frac{d}{dr} e^{ik \cdot r} = i \frac{\mathbf{k} \cdot \mathbf{r}}{r} e^{ik \cdot r}. \quad (4.129)$$

Finally, from Eq. (4.123), we have

$$\delta_s = \delta_r - \frac{du}{dr} = \delta_r + \beta \int \delta_{rk} e^{ik \cdot r} \left(\frac{\mathbf{k} \cdot \mathbf{r}}{kr} \right)^2 d^3 k^*, \quad (4.130)$$

where we have neglected the second term in (4.123) because it is negligible for large r .

Now, we can multiply Eq. (4.130) on both sides by $V^{-1} e^{-ik' \cdot r} d^3 r$ and integrate it. We then obtain from Eq. (3.14) the Fourier transform of (4.130) as

$$\delta_{sk} = \delta_{rk} + \beta \int \delta_{rk'} I(k, k') d^3 k', \quad (4.131)$$

(we switched k, k') where

$$I(k, k') = (2\pi)^{-3} \int e^{i(k' - k) \cdot r} \left(\frac{\mathbf{k}' \cdot \mathbf{r}}{k' r} \right)^2 d^3 r. \quad (4.132)$$

The redshift distortion then introduces a mode–mode coupling. This coupling can be broken in the useful limit of surveys of very small angular scales. In fact, if we can assume that the cosine

$$\mu = \frac{\mathbf{k} \cdot \mathbf{r}}{kr} \quad (4.133)$$

is almost constant (that is, the survey spans a small solid angle), then we have $I(k, k') = \mu^2 \delta_D(\mathbf{k}' - \mathbf{k})$ and

$$\delta_{sk} = \delta_{rk}(1 + \beta\mu^2). \quad (4.134)$$

The power spectrum defined in Eq. (3.19) reads

$$P_s(\mathbf{k}) = V \delta_{rk}^2 (1 + \beta\mu^2)^2 = P_r(\mathbf{k})(1 + \beta\mu^2)^2. \quad (4.135)$$

If we average it over angles, we get

$$P_s(k) = P_r(k)(1 + 2\beta\langle\mu^2\rangle + \beta^2\langle\mu^4\rangle), \quad (4.136)$$

where the average $\langle f(\mu) \rangle = (4\pi)^{-1} \int_0^\pi f(\mu) \sin\theta d\theta \int_0^{2\pi} d\phi = (1/2) \int_{-1}^1 f(\mu) d\mu$ gives

$$\langle\mu^2\rangle = \frac{1}{2} \int_{-1}^1 \mu^2 d\mu = \frac{1}{3}, \quad \langle\mu^4\rangle = \frac{1}{2} \int_{-1}^1 \mu^4 d\mu = \frac{1}{5}. \quad (4.137)$$

We then obtain [98]

$$P_s(k) = P_r(k)(1 + 2\beta/3 + \beta^2/5). \quad (4.138)$$

The power spectrum is then boosted in redshift space, because the velocity field is directed toward mass concentrations. As a result, galaxies seem more concentrated when seen in redshift space. Remember that we have dropped the subscript g but both P_r and P_s refer to the galaxies.

As we anticipated, at very small scales, on the other hand, the velocity orientation can be assumed to be random. Then, on these scales, the net effect is that galaxies in redshift space seem distributed over a larger volume and the power spectrum is therefore decreased. Empirical studies have shown that an approximation valid both at small and large scales is given by the following correction [99]:

$$P_s(k) = P_r(k)G(\beta, y), \quad (4.139)$$

where

$$G(\beta, y) = \frac{\pi^{1/2} \operatorname{erf}(y)}{8 y^5} (3\beta^2 + 4\beta y^2 + 4y^4) - \frac{e^{-y^2}}{4y^4} [\beta^2(3 + 2y^2) + 4\beta y^2], \quad (4.140)$$

and $y = k\sigma_v H_0^{-1}$ with σ_v being the cloud velocity dispersion along the line of sight. On small scales the effect is to change the slope by a factor k^{-1} , whereas on large scales the effect is to raise the amplitude as in Eq. (4.138).

4.9 Baryons, photons, and neutrinos

So far we have considered only perfect fluids which do not possess anisotropic shear (i.e. $T_j^i = 0$) and do not interact except gravitationally. Because of this,

their perturbation dynamics has been completely described by functions of the wavenumber modulus k alone, the density contrast, and the gradient of velocity field.

However, the cosmic mix contains imperfect fluids for which such a description is insufficient. When the interaction terms or the energy-momentum tensor depend on the full momentum vector $\mathbf{P}$, the fluid needs to be described by its full distribution function $f(\mathbf{P}, \mathbf{x}, t)$. Radiation, baryons, and neutrinos are such fluids and in the following we discuss their full perturbation equations. This section is only meant to summarize the main results that will be used elsewhere. We urge the reader to familiarize themselves with the full treatment in standard cosmology textbooks (e.g., [74], whose notation we follow closely).

We shall take the perturbed metric (4.13) with the cosmic time $t = \int a d\eta$, i.e.

$$ds^2 = -(1 + 2\Psi)dt^2 + a^2(t)(1 + 2\Phi)\delta_{ij}dx^i dx^j. \quad (4.141)$$

Given a distribution function $f(\mathbf{P}, \mathbf{x}, t)$, the energy-momentum tensor for a fluid in the full general relativistic framework is

$$T^\mu_\nu(\mathbf{x}, t) = \frac{g_i}{(2\pi)^3} \int dP_1 dP_2 dP_3 \sqrt{-g} \frac{P^\mu P_\nu}{P^0} f(\mathbf{P}, \mathbf{x}, t), \quad (4.142)$$

where $P^0 \equiv dt/d\lambda_s$, $P^i \equiv dx^i/d\lambda_s$ (λ_s is a parameter that characterizes the particle's path), and g_i are the internal degrees of freedom. For a particle of mass m one has $P^\mu P_\mu = -m^2$. From the expression (4.142) we can derive the usual energy density $-T^0_0$ and pressure $T^i_i/3$. We see that the term that contributes to the shear T^i_j is proportional to

$$\frac{P^i P_j}{(P^0)^2} = \frac{dx^i dx_j}{dt^2} = v^i v_j, \quad (4.143)$$

which is a second-order quantity and therefore is negligible in linear perturbation theory. For massless particles, however, this is not the case. Photons and neutrinos (massless or massive but relativistic) therefore contribute to the shear term. Because of the mass constraint on P^μ there are only three momentum degrees of freedom. We can choose the spatial magnitude

$$P^2 \equiv g_{ij} P^i P^j, \quad (4.144)$$

and the unit direction vector

$$\hat{P}^i \equiv P^i / |P|, \quad (4.145)$$

such that $\delta_{ij} \hat{P}^i \hat{P}^j = 1$.

Plugging $P^i = |P| \hat{P}^i$ into Eq. (4.144), we find

$$P^2 = a^2(1 + 2\Phi)(\delta_{ij} \hat{P}^i \hat{P}^j) P^2 = a^2(1 + 2\Phi) P^2, \quad (4.146)$$

which gives $|P| = p(1 - \Phi)/a$ at first-order. Hence the spatial vector P^i can be written as

$$P^i = \frac{1 - \Phi}{a} p \hat{p}^i. \quad (4.147)$$

The photons satisfy the relation $g_{\mu\nu} P^\mu P^\nu = 0$, which translates into the following condition

$$-(1 + 2\Psi)(P^0)^2 + p^2 = 0. \quad (4.148)$$

We then obtain the time-component of P^μ :

$$P^0 = p(1 - \Psi). \quad (4.149)$$

The process of collisions between particles can be described by the Boltzmann equation

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x^i} \frac{dx^i}{dt} + \frac{\partial f}{\partial p} \frac{dp}{dt} + \frac{\partial f}{\partial \hat{p}^i} \frac{d\hat{p}^i}{dt} \quad (4.150)$$

$$= C[f], \quad (4.151)$$

where $f(p, \hat{p}^i, x^i, t)$ is the distribution function and $C[f]$ describes a collision term. The r.h.s. of Eq. (4.150) should be evaluated up to the first-order for the line element (4.141). The last term in Eq. (4.150) vanishes at the linear level, since both $\partial f / \partial \hat{p}^i$ and $d\hat{p}^i / dt$ are first-order terms. The collision term in Eq. (4.151) is different depending on the kind of matter species. For photons we need to compute it for the process of Compton interaction between protons and electrons (i.e. baryons). Dark matter interacts very weakly with other particles so that the collision term is set to zero.

Let us consider the Boltzmann equation (4.151) for the photons. From Eqs. (4.147) and (4.149) it follows that

$$\frac{dx^i}{dt} = \frac{P^i}{P^0} = \frac{1 - \Phi + \Psi}{a} \hat{p}^i. \quad (4.152)$$

From the time component of the geodesic equation $dP^0/d\lambda_s = -\Gamma_{\mu\nu}^0 P^\mu P^\nu$, we obtain the following relation [74]

$$\frac{dp}{dt} = -p \left(H + \frac{\partial \Phi}{\partial t} + \frac{\hat{p}^i}{a} \frac{\partial \Psi}{\partial x^i} \right). \quad (4.153)$$

Then the l.h.s. of the Boltzmann equation (4.151) yields

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\hat{p}^i}{a} \frac{\partial f}{\partial x^i} - p \frac{\partial f}{\partial p} \left(H + \frac{\partial \Phi}{\partial t} + \frac{\hat{p}^i}{a} \frac{\partial \Psi}{\partial x^i} \right). \quad (4.154)$$

Note that we have neglected the product of $\partial f / \partial x^i$ and $\Phi(\Psi)$ because $\partial f / \partial x^i$ is a first-order term (f does not depend on x^i at zero-th order).

Recall that in the unperturbed background the photons with the temperature T obey the Bose–Einstein distribution function

$$f^{(0)}(t, p) = [\exp(p/T) - 1]^{-1}, \quad (4.155)$$

where we have neglected the chemical potential μ . Note that in this background the temperature T depends on the time t only: $T \propto 1/a(t)$. In the perturbed Universe we can define the temperature perturbation:

$$\Theta(t, \mathbf{x}, \hat{p}^i) \equiv \delta T / T, \quad (4.156)$$

which is chosen as a perturbation variable of photons instead of $\delta_\gamma = 4\delta T / T$ (recall that $\rho_r \propto \rho_\gamma \propto T^4$). We assume that Θ does not depend on the magnitude p since in a Compton scattering this is approximately conserved. The distribution function is given by

$$f(t, p, \mathbf{x}, \hat{p}) = \left\{ \exp \left[\frac{p}{T(t)[1 + \Theta(t, \mathbf{x}, \hat{p}^i)]} \right] - 1 \right\}^{-1}. \quad (4.157)$$

If $\Theta \ll 1$, one can expand this about the background value $f^{(0)}$ by using the relation $T \partial f^{(0)} / \partial T = -p \partial f^{(0)} / \partial p$:

$$f = f^{(0)} - p \frac{\partial f^{(0)}}{\partial p} \Theta, \quad (4.158)$$

which is valid at first-order. Plugging Eq. (4.158) into Eq. (4.154) and collecting the first-order terms, we obtain

$$\frac{df^{(1)}}{dt} = -p \frac{\partial f^{(0)}}{\partial p} \left[\frac{\partial \Theta}{\partial t} + \frac{\hat{p}^i}{a} \frac{\partial \Theta}{\partial x^i} + \frac{\partial \Phi}{\partial t} + \frac{\hat{p}^i}{a} \frac{\partial \Psi}{\partial x^i} \right]. \quad (4.159)$$

In Fourier space, we expand Θ as

$$\Theta(\mathbf{r}) = \frac{1}{(2\pi)^3} \int d^3k \Theta_k e^{i\mathbf{k} \cdot \mathbf{r}}, \quad (4.160)$$

where the time dependence on the l.h.s. is implicit. Instead of the unit vector $\hat{\mathbf{p}}$, we can use the direction cosine

$$\mu = \frac{\mathbf{k} \cdot \hat{\mathbf{p}}}{k}, \quad (4.161)$$

where $k = |\mathbf{k}|$ is the magnitude of the wavevector $\mathbf{k}$.

It turns out convenient to integrate out the angular dependence of $\Theta(k)$ by defining the multipoles

$$\Theta_\ell \equiv \frac{1}{(-i)^\ell} \int_{-1}^1 \frac{d\mu}{2} \mathcal{P}_\ell(\mu) \Theta(\mu), \quad (4.162)$$

where $\mathcal{P}_\ell$ is the Legendre polynomial of order ℓ . The first polynomials are $\mathcal{P}_0(\mu) = 1$, $\mathcal{P}_1(\mu) = \mu$, and $\mathcal{P}_2(\mu) = (3\mu^2 - 1)/2$. So $\ell = 0$ defines the monopole Θ_0 , which is simply the angular average of Θ , while $\ell = 1$ gives the dipole term.

The next step is to evaluate the collision term $C[f]$ for the photons, determined mainly by the Compton scattering process $e^- + \gamma \leftrightarrow e^- + \gamma$. The interested reader may wish to consult the textbook [74], since its complete derivation requires several pages. The final result is however pretty simple:

$$C[f] = -p \frac{\partial f^{(0)}}{\partial p} n_e \sigma_T [\Theta_0 - \Theta(\hat{p}) + \hat{p} \cdot \mathbf{v}_b], \quad (4.163)$$

where n_e is the electron density, σ_T is the Thomson cross section, and $\mathbf{v}_b$ is the velocity of the electrons (here classified as usual in cosmology among the “baryons”). In the absence of $\mathbf{v}_b$, a strong coupling (i.e. the limit of large $n_e \sigma_T$) means that $\Theta \approx \Theta_0$, i.e. that the perturbation is fully described by its monopole term.

Equating (4.159) and (4.163) we obtain

$$\frac{\partial \Theta}{\partial t} + \frac{\hat{p}^i}{a} \frac{\partial \Theta}{\partial x^i} + \frac{\partial \Phi}{\partial t} + \frac{\hat{p}^i}{a} \frac{\partial \Psi}{\partial x^i} = n_e \sigma_T [\Theta_0 - \Theta(\hat{p}) + \hat{p} \cdot \mathbf{v}_b]. \quad (4.164)$$

In Fourier space we have that $\partial \Theta / \partial x^j \rightarrow i k_j \Theta$, and as usual we omit the subscript k for Fourier-transformed quantities. Moreover, since we always assume the fluid to be irrotational, $\mathbf{v}_b$ is directed along $\mathbf{k}$ so that

$$v_b^i = v_b k^i / k, \quad (4.165)$$

where v_b is the velocity modulus. Then Eq. (4.164) can be written as

$$\frac{\partial \Theta}{\partial t} + \frac{i k \mu}{a} \Theta + \frac{\partial \Phi}{\partial t} + \frac{i k \mu}{a} \Psi = n_e \sigma_T [\Theta_0 - \Theta(\hat{p}) + \mu v_b]. \quad (4.166)$$

In terms of the conformal time η , this equation reduces to

$$\Theta' + i k \mu \Theta + \Phi' + i k \mu \Psi = -\tau'_{\text{op}} [\Theta_0 - \Theta(\hat{p}) + \mu v_b], \quad (4.167)$$

where we have introduced the *optical depth* τ_{op} defined as

$$\tau_{\text{op}} \equiv \int_{\eta}^{\eta_0} n_e \sigma_T a \, d\tilde{\eta}. \quad (4.168)$$

Note that a prime represents a derivative with respect to η (with η_0 being the conformal time today). The optical depth is small when the free electron density n_e is small, i.e. at late times. In the limit $\tau_{\text{op}} \rightarrow 0$, the photon fluctuations decouple from the baryons. In this limit, i.e. putting the r.h.s. of Eq. (4.167) to zero, the equation can be applied as well to massless neutrinos. On the contrary, when n_e is large such that $\tau_{\text{op}} \gg 1$, i.e. before the recombination epoch, the photons and baryons are tightly coupled.

Now we should redo the calculations for massive particles, in order to derive the Boltzmann equation for baryons. Instead of the momentum magnitude p , we introduce the energy $E = \sqrt{p^2 + m^2}$. For non-relativistic matter, all terms of order $v^2 = (p/E)^2$ or higher can be neglected. This allows us to write the Boltzmann equations for the distribution f_m in terms of the first and second moments, i.e. of the number density and velocity

$$n_m \equiv \int \frac{d^3 p}{(2\pi)^3} f_m, \quad v_m^i \equiv \frac{1}{n_m} \int \frac{d^3 p}{(2\pi)^3} \frac{p \hat{p}^i}{E} f_m. \quad (4.169)$$

Recall that in Section 4.3 we have derived the set of perfect fluid equations (4.48) and (4.51) for the pressureless matter component. For the baryons, however, there is again a collisional term on the r.h.s. of the Boltzmann equation. This term contributes only to the velocity equation (4.51), i.e. the first moment of the baryon distribution f_b , since both Compton scattering (that couples electrons and photons) and Coulomb scattering (that couples electrons and protons) conserve the number of “baryons.” Therefore the velocity equation acquires an additional term coming from the scattering process. In Fourier space the full set of coupled photon–baryon equations reads

$$\Theta' = -\Phi' - ik\mu(\Theta + \Psi) - \tau'_{\text{op}}(\Theta_0 - \Theta + \mu v_b), \quad (4.170)$$

$$\delta'_b = -ikv_b - 3\Phi', \quad (4.171)$$

$$v'_b = -\mathcal{H}v_b - ik\Psi + \frac{\tau'_{\text{op}}}{R_s}(3i\Theta_1 + v_b), \quad (4.172)$$

where

$$R_s \equiv \frac{3}{4} \frac{\rho_b}{\rho_\gamma}. \quad (4.173)$$

Note that we have used the relation $\theta_b = ik_j v_b^j = ikv_b$ to convert θ_b to v_b in Eqs. (4.171) and (4.172).

The cold dark matter does not have an electromagnetic interaction so that the perturbations are not affected by the collision term unlike the baryons. Hence the

CDM perturbations δ_c and v_c obey the following equations

$$\delta'_c = -ikv_c - 3\Phi', \quad (4.174)$$

$$v'_c = -\mathcal{H}v_c - ik\Psi, \quad (4.175)$$

which are of course identical to the matter equations we derived in Eqs. (4.86) and (4.87) (with $\Psi = -\Phi$).

Now we can derive an equation for the monopole term Θ_0 that will be useful later on. Multiplying $\mathcal{P}_0(\mu)$ and $\mathcal{P}_1(\mu)$ for Eq. (4.170) and integrating them in terms of μ in the range $[-1, 1]$, we obtain

$$\Theta'_0 + k\Theta_1 = -\Phi', \quad (4.176)$$

$$\Theta'_1 - \frac{k}{3}(\Theta_0 + \Psi) = \tau'_{\text{op}} \left(\Theta_1 - \frac{i}{3}v_b \right), \quad (4.177)$$

where we have neglected the term Θ_2 (see problem 4.2). In the tight coupling regime ($\tau_{\text{op}} \gg 1$) one has $|\tau'_{\text{op}}\Theta_1| \gg |\Theta'_1|$ in Eq. (4.177). Then the second term, $(k/3)\Theta_0$, in Eq. (4.177) is the same order as the r.h.s., giving $\Theta_1 \sim k\Theta_0/\tau'_{\text{op}} \sim (k\eta/\tau_{\text{op}})\Theta_0$. In the tight coupling limit ($\tau_{\text{op}} \gg 1$) this means that $\Theta_1 \ll \Theta_0$ for the modes around the Hubble radius ($k\eta \sim 1$). This property generally holds for the multipole moments Θ_ℓ , i.e. $\Theta_{\ell+1}/\Theta_\ell \sim k\eta/\tau_{\text{op}}$. Hence it is a good approximation to neglect the moments Θ_ℓ ($\ell \geq 2$) compared to monopole and dipole moments in the tight coupling regime.

If we rewrite Eq. (4.172) in the form $v_b = -3i\Theta_1 + (R_s/\tau'_{\text{op}})[v'_b + \mathcal{H}v_b + ik\Psi]$, then the term $(R_s/\tau'_{\text{op}})[v'_b + \mathcal{H}v_b + ik\Psi]$ is suppressed relative to others by a factor of $1/\tau_{\text{op}}$, giving the relation $v_b \simeq -3i\Theta_1$ at the lowest order. Substituting this relation into Eq. (4.172), we get

$$v_b \simeq -3i\Theta_1 - 3i\frac{R_s}{\tau'_{\text{op}}} \left(\Theta'_1 + \mathcal{H}\Theta_1 - \frac{k}{3}\Psi \right). \quad (4.178)$$

Plugging Eq. (4.178) into Eq. (4.177) and taking the η derivative of Eq. (4.176) to eliminate the term Ψ' , we find (see the problem 4.3)

$$\Theta''_0 + \frac{R_s}{1+R_s}\mathcal{H}\Theta'_0 + k^2c_s^2\Theta_0 = -\frac{k^2}{3}\Psi - \frac{R_s}{1+R_s}\mathcal{H}\Phi' - \Phi'', \quad (4.179)$$

where

$$c_s^2 \equiv \frac{\delta P_\gamma}{\delta\rho_\gamma + \delta\rho_b} = \frac{1}{3(1+R_s)} \quad (4.180)$$

is the effective sound speed squared of the coupled baryon–photon plasma. Recall that R_s is defined in Eq. (4.173). The last equality in Eq. (4.180) holds by using the relations $\delta P_\gamma = \delta\rho_\gamma/3$ and $\delta\rho_\gamma/\rho_\gamma = (4/3)\delta\rho_b/\rho_b$ (the second one comes from

$\rho_\gamma \propto a^{-4} \propto T^4$ and $\rho_b \propto a^{-3} \propto T^3$). In the absence of baryons the sound speed reduces to that for a relativistic fluid: $c_s = c/\sqrt{3}$, where we have recovered the speed of light c . When baryons are present, the sound speed decreases because the fluid becomes heavier.

Equation (4.179) is very convenient because it shows how the direction-averaged temperature anisotropy Θ_0 responds to gravity, i.e. to Ψ and Φ . In Section 5.3 we shall discuss the solution of this equation to confront the predicted temperature anisotropies with CMB observations.

The gravitational potentials Φ and Ψ are determined by the combined action of all the matter fields, baryons (b), cold dark matter (c), photons (γ), neutrinos (ν), and dark energy (DE). If we neglect the contribution of dark energy, the (00) component of the Einstein equation (4.57) reads

$$k^2 \Phi + 3\mathcal{H}(\Phi' - \mathcal{H}\Psi) = 4\pi G a^2 (\rho_m \delta_m + 4\rho_r \Theta_{r,0}) , \quad (4.181)$$

where

$$\rho_m \delta_m \equiv \rho_b \delta_b + \rho_c \delta_c , \quad (4.182)$$

$$\rho_r \Theta_{r,i} \equiv \rho_\gamma \Theta_i + \rho_\nu \mathcal{N}_i , \quad (i = 0, 1, 2, \dots). \quad (4.183)$$

Here Θ_0 and $\mathcal{N}_0$ are the monopoles of photons and massless neutrinos, respectively, which are related to the energy momentum tensors via $^{(\gamma)}T_0^0 = -\rho_\gamma(1 + 4\Theta_0)$ and $^{(\nu)}T_0^0 = -\rho_\nu(1 + 4\mathcal{N}_0)$. From Eqs. (4.58) and (4.181) it follows that

$$k^2 \Phi = 4\pi G a^2 \left[\rho_m \delta_m + 4\rho_r \Theta_{r,0} + \frac{3\mathcal{H}}{k} (i\rho_m v_m + 4\rho_r \Theta_{r,1}) \right] , \quad (4.184)$$

where we have used the fact that the velocity v_r for radiation is related to the dipole $\Theta_{r,1}$ via $v_r = -3i\Theta_{r,1}$ [74].

There is also another relation that relates the sum of the gravitational potentials Φ and Ψ with the anisotropic stresses of photons and neutrinos:

$$k^2(\Phi + \Psi) = -32\pi G a^2 \rho_r \Theta_{r,2} . \quad (4.185)$$

If the quadrupole $\Theta_{r,2}$ is negligibly small, it follows that $\Phi \simeq -\Psi$. Equation (4.185) assumes a particularly simple form during the radiation-dominated era and for super-horizon scales. During the radiation-dominated era we can use the Friedmann equation (2.17) in flat space to rewrite it as

$$k^2(\Phi + \Psi) = -12\mathcal{H}^2 f_\nu \mathcal{N}_2 , \quad (4.186)$$

where

$$f_\nu = \frac{\rho_\nu}{\rho_\nu + \rho_\gamma} . \quad (4.187)$$

In fact, the dominant term in the quadrupole $\Theta_{r,2}$ is the neutrinos since the photon–baryon plasma behaves almost as a perfect fluid. Note that $f_v \approx \Omega_v \approx \text{constant}$ during the radiation-dominated era but not afterward. For adiabatic initial conditions the large-scale fluctuations in radiation equal the fluctuations in the neutrino component since they are both relativistic. Therefore we can put $\Theta_{r,0} = \mathcal{N}_0$. Now from Eq. (4.181), assuming $k \ll \mathcal{H}$ and constant Φ (i.e. discarding the first and second term on the l.h.s.) and negligible ρ_m , we obtain

$$3\mathcal{H}^2\Psi = -16\pi G a^2 \rho_r \Theta_{r,0}. \quad (4.188)$$

It then follows that

$$\Psi = -2\Theta_{r,0} = -2\mathcal{N}_0. \quad (4.189)$$

This replaces Eq. (4.71) when we cannot assume $\Phi = -\Psi$. For the neutrinos we can use Eqs. (4.176) and (4.177) with $\mathcal{N}_{0,1}$ instead of $\Theta_{0,1}$ and no coupling to baryons. However the quadrupole $\mathcal{N}_2$ that we neglected in Eq. (4.177) is no longer negligible. Multiplying Eq. (4.170) by $\mathcal{P}_2(\mu)$ and integrating it with respect to μ , we obtain an equation for $\mathcal{N}_2$. Neglecting $\mathcal{N}_3$ for $k \ll \mathcal{H}$ and eliminating $\mathcal{N}_1$, one finds

$$\mathcal{N}_2'' = \frac{2k^2}{15}(\Psi + \mathcal{N}_0 - 2\mathcal{N}_2). \quad (4.190)$$

Equation (4.186) can be written as

$$\mathcal{N}_2 = A\mathcal{H}^{-2}, \quad (4.191)$$

where $A = -k^2(\Phi + \Psi)/(12f_v)$ is approximately constant since the potentials are constant on super-horizon scales. Differentiating Eq. (4.191) twice with respect to η (notice that $\mathcal{H}'/\mathcal{H}^2 \simeq -1 = \text{constant}$ in this regime) we obtain

$$\mathcal{N}_2''' = 2A = -k^2(\Phi + \Psi)/(6f_v). \quad (4.192)$$

Combining Eqs. (4.189), (4.190), (4.192) and taking the limit $k \ll \mathcal{H}$, it follows that

$$\Phi = -\Psi \left(1 + \frac{2}{5}f_v\right). \quad (4.193)$$

This sets the initial condition for the relation between the potentials at early times.

4.10 The matter power spectrum

The large-scale structure of the Universe started to grow after the epoch of the radiation–matter equality. Since non-relativistic matter has a negligible pressure relative to its energy density, the gravitational attraction becomes stronger than the

pressure repulsion in the matter-dominated epoch. The perturbations of pressureless matter, especially the CDM perturbations, are responsible for the formation of galaxies. We can quantify the matter distribution in the Universe by measuring the correlation function or the power spectrum of the galaxies we observe in the sky.

In order to derive the spectrum of matter perturbations today, we need to know the evolution of $\Phi(k, t)$ from the early Universe to the present epoch. The cosmic inflation in the early Universe that is needed to solve the problems of flatness and horizon is believed to have generated the seeds of anisotropies through the quantum fluctuations of a scalar field (called “inflaton”) [70].

In the simplest version of the single-field inflation scenario the perturbations are “frozen” after the scale $\lambda_p = (2\pi/k)a$ of perturbations leaves the Hubble radius (H^{-1}) during the accelerated expansion (we are considering the perturbations whose wavelengths are smaller than the Hubble radius at the onset of inflation). The perturbations in the super-horizon regime (characterized by $k \lesssim \mathcal{H} = aH$) evolve only adiabatically after the first Hubble radius crossing ($k = \mathcal{H}$). After inflation the comoving Hubble radius, $\mathcal{H}^{-1}$, begins to increase as long as the Universe undergoes a decelerated expansion ($a \propto t^p$ with $p < 1$). During radiation and matter eras, the perturbations cross inside the Hubble radius again ($k \gtrsim \mathcal{H}$). The epoch of this second Hubble radius crossing ($k = \mathcal{H}$) depends upon the wavelength of perturbations. The crossing occurs earlier for smaller scale perturbations.

During inflation nearly scale-invariant density perturbations are generated from the quantum fluctuation of a scalar field with a potential, which sets up initial conditions for the gravitational potential. Inflation generally predicts nearly scale-invariant density perturbations, which are consistent with the CMB anisotropies observed by COBE and WMAP. The initial power spectrum of Φ generated during inflation is given by

$$P_{\Phi}^{(i)} \equiv \langle |\Phi(k, a_i)|^2 \rangle = \frac{50\pi^2}{9k^3} \left(\frac{k}{H_0} \right)^{n_s-1} \delta_H^2, \quad (4.194)$$

(conventionally there is no factor of volume in this definition of power spectrum) where n_s is the spectral index and δ_H^2 represents the amplitude of the gravitational potential (see Refs. [70, 74, 94] for the derivation of Eq. (4.194)). The WMAP 5-year observations constrain these values to be $n_s = 0.960^{+0.014}_{-0.013}$ and $\delta_H^2 \simeq 3.2 \times 10^{-10}$ [15]. Note that $n_s = 1$ corresponds to the scale-invariant spectrum with $k^3 \langle |\Phi(k, a_i)|^2 \rangle = \text{constant}$.¹

¹ Although in mathematics “scale-invariant” means generally a power-law behavior, cosmologists use this term to denote a flat spectrum k^0 . In the context of inflation the power spectrum is usually defined by $\hat{P}_{\Phi}^{(i)} \equiv k^3/(2\pi^2) \langle |\Phi(k, a_i)|^2 \rangle$, but we adopt the definition (4.194) to match with the standard notation used to evaluate the matter power spectrum.

In order to obtain the gravitational potential today, we need to solve the equation for $\Phi(k, t)$ from the beginning of the radiation-dominated epoch to the present. The evolution of the gravitational potential during the early cosmological epoch depends on the modes k . The wavenumber k_{eq} that characterizes the border of “large-scale” and “small-scale” modes corresponds to the one that entered the Hubble radius at the radiation–matter equality, i.e. $k_{\text{eq}} = a_{\text{eq}} H(a_{\text{eq}})$. Using Eq. (2.78) together with $H(a_{\text{eq}})/H_0 = [2\Omega_m^{(0)}/a_{\text{eq}}^3]^{1/2}$, we find

$$k_{\text{eq}} = H_0 \sqrt{\frac{2\Omega_m^{(0)}}{a_{\text{eq}}}} = 0.073 \Omega_m^{(0)} h^2 \text{ Mpc}^{-1}. \quad (4.195)$$

Let us first consider the large-scale modes with $k \ll k_{\text{eq}}$. In Section 4.4 we have already seen that there is a solution with $\Phi = \text{constant}$ for a single-fluid system. However it remains to see what happens during the transition from the radiation era to the matter era. In the following we shall neglect the contribution of baryons as well as the quadrupole $\Theta_{r,2}$. Under the super-horizon approximation ($k \ll \mathcal{H}$), we obtain the following approximate equations from Eqs. (4.181), (4.174), and (4.176):

$$3\mathcal{H}(\Phi' + \mathcal{H}\Phi) = 4\pi G a^2 (\rho_c \delta_c + 4\rho_r \Theta_{r,0}), \quad (4.196)$$

$$\delta'_c = -3\Phi', \quad (4.197)$$

$$\Theta'_{r,0} = -\Phi'. \quad (4.198)$$

In Eqs. (4.196) and (4.198), $\Theta_{r,0}$ includes the contributions of both radiation and neutrinos. From Eqs. (4.197) and (4.198) it follows that $\delta_c = 3\Theta_{r,0} + C$, where C is an integration constant. As we have already mentioned in the previous section, the radiation perturbations $\delta_r = \delta\rho_r/\rho_r$ are related to the matter perturbation $\delta_m = \delta\rho_m/\rho_m$ via $\delta_r = (4/3)\delta_m$ by noting that $\rho_r \propto T^4$ and $\rho_m \propto T^3$. This translates into $\delta_m = 3\Theta_{r,0}$, which is called the adiabatic condition (applied to both dark matter and baryons). This shows that $C = 0$ for CDM as long as the adiabatic initial condition is respected.

Plugging $\Theta_{r,0} = \delta_c/3$ into Eq. (4.196) and introducing the variable $y \equiv a/a_{\text{eq}} = \rho_c/\rho_r$, we find

$$y \frac{d\Phi}{dy} + \Phi = \frac{3y+4}{6(y+1)} \delta_c, \quad (4.199)$$

where we have used the relations $dy/d\eta = \mathcal{H}y$ and $3H^2 = 8\pi G\rho_c(1 + 1/y)$. Taking the y -derivative of Eq. (4.199) and using Eq. (4.197), we obtain

$$\frac{d^2\Phi}{dy^2} + \frac{21y^2 + 54y + 32}{2y(y+1)(3y+4)} \frac{d\Phi}{dy} + \frac{1}{y(y+1)(3y+4)} \Phi = 0. \quad (4.200)$$

This has the following analytic solution [90]

$$\Phi(y) = c_1 \frac{\sqrt{y+1}}{y^3} + c_2 \frac{9y^3 + 2y^2 - 8y - 16}{y^3}, \quad (4.201)$$

where the integration constants c_1 and c_2 are determined by imposing the initial conditions $\Phi_i = \Phi(0)$ and $(d\Phi/dy)_i = 0$. This gives $c_1 = 16c_2 = (8/5)\Phi(0)$, so that the solution is

$$\Phi(y) = \Phi(0) \frac{9y^3 + 2y^2 - 8y - 16 + 16\sqrt{y+1}}{10y^3}. \quad (4.202)$$

In the limit that $y = a/a_{\text{eq}} \rightarrow \infty$, the gravitational potential approaches $\Phi \rightarrow (9/10)\Phi(0)$. Hence, for super-horizon perturbations, the gravitational potential decreases by 10% during the transition from the radiation era to the matter era.

Let us next consider the evolution of Φ during the deep radiation-dominated epoch. Since $c_s^2 \simeq 1/3$, $\mathcal{H}' \simeq -\mathcal{H}^2$, and $\mathcal{H} \simeq 1/\eta$ in this regime, Eq. (4.66) gives

$$\Phi'' + \frac{4}{\eta}\Phi' + \frac{k^2}{3}\Phi = 0. \quad (4.203)$$

The solution of this equation satisfying the initial conditions $\Phi = \Phi_I$ and $d\Phi/d\eta = 0$ at $\eta = 0$ is

$$\Phi(k, \eta) = 3\Phi_I \frac{\sin(k\eta/\sqrt{3}) - (k\eta/\sqrt{3})\cos(k\eta/\sqrt{3})}{(k\eta/\sqrt{3})^3}. \quad (4.204)$$

For the modes outside the Hubble radius ($k \ll \mathcal{H}$, i.e. $k\eta \ll 1$) we have $\Phi(k, \eta) \simeq \Phi_I [1 - (k\eta)^2/10]$, which means that the gravitational potential is nearly constant, as we know already. The modes with $k \gg k_{\text{eq}}$ crossed inside the Hubble radius ($k\eta > 1$) before the radiation–matter equality. These small-scale perturbations started to decay after the Hubble radius crossing. Asymptotically ($k\eta \gg 1$) the gravitational potential decreases as $1/(k\eta)^2$ with oscillations (during the radiation era). For larger k this decay started earlier, which implies that the amplitude of the resulting gravitational potential is suppressed for perturbations on smaller scales. After the Universe enters the matter-dominated epoch, the amplitude of Φ approaches a constant value.

Thus we have shown that the evolution of the gravitational potential depends on the scales of perturbations. In order to describe its evolution for each wavenumber k during the transfer epoch (from the radiation era to the epoch at $a = a_T$), we introduce the *transfer function*

$$T(k) \equiv \frac{\Phi(k, a_T)}{\Phi_{\text{LS}}(k, a_T)}, \quad (4.205)$$

where $\Phi_{\text{LS}}(k, a_T)$ is the large-scale solution decreased by an amount 9/10 compared to the primordial value $\Phi(k, a_i)$ generated during inflation:

$$\Phi_{\text{LS}}(k, a_T) = \frac{9}{10} \Phi(k, a_i). \quad (4.206)$$

The typical value of a_T is around $a_T \sim 0.03$. Note that for $a > a_T$ the evolution of Φ becomes independent of k (i.e. $\Phi = \text{constant}$ during the matter era).

In general, the transfer function has to be derived numerically by integrating the equations for each k mode. A popular fit has been given by Bardeen, Bond, Kaiser, and Szalay (BBKS) [100]:

$$T(x) = \frac{\ln(1 + 0.171x)}{0.171x} \left[1 + 0.284x + (1.18x)^2 + (0.399x)^3 + (0.490x)^4 \right]^{-1/4}, \quad (4.207)$$

where $x \equiv k/k_{\text{eq}}$. On large scales characterized by the condition $x = k/k_{\text{eq}} \ll 1$ the BBKS transfer function reduces to $T(x) \simeq 1$, which means that $\Phi(k, a_T) = (9/10)\Phi(k, a_i)$ as expected. On small scales with $x = k/k_{\text{eq}} \gg 1$ the transfer function has a k -dependence $T(k) \propto (\ln k)/k^2$, which means that the gravitational potential $\Phi(k, a_T)$ is suppressed for increasing k . Note that the logarithmic correction comes from matching two solutions of dark matter perturbations during the radiation and matter eras.

When $a > a_T$ the gravitational potential Φ remains constant during the matter era at least in standard General Relativity, but after the Universe has entered the epoch of cosmic acceleration Φ is expected to vary. In order to quantify this, we introduce the growth function $D(a)$:

$$\frac{\Phi(a)}{\Phi(a_T)} = \frac{D(a)}{a} \quad (a > a_T). \quad (4.208)$$

In Fig. 4.1 we illustrate the evolution of the quantity $D(a)/a$ in both CDM and Λ CDM models. In the CDM model the gravitational potential is constant for $a > a_T$ and hence $D(a) = a$. In the Λ CDM model, the constant evolution of $\Phi(a)$ is followed by its decay around the end of the matter era. This is associated with the decreases of both Ω_m and the growth rate of δ_m . This variation of Φ leads to the integrated Sachs–Wolfe (ISW) effect in the CMB temperature anisotropies, as we will see in more detail in Sections 4.11 and 5.3.

Combining Eqs. (4.205), (4.206), and (4.208), we find that the gravitational potential today is given by

$$\Phi(k, a_0) = \frac{9}{10} \Phi(k, a_i) T(k) D(a_0), \quad (4.209)$$

where we have used $a_0 = 1$.

After the Universe enters the matter era, the radiation perturbation $\Theta_{r,0}$ is negligible relative to the matter perturbation δ_m . For the Λ CDM model or the models

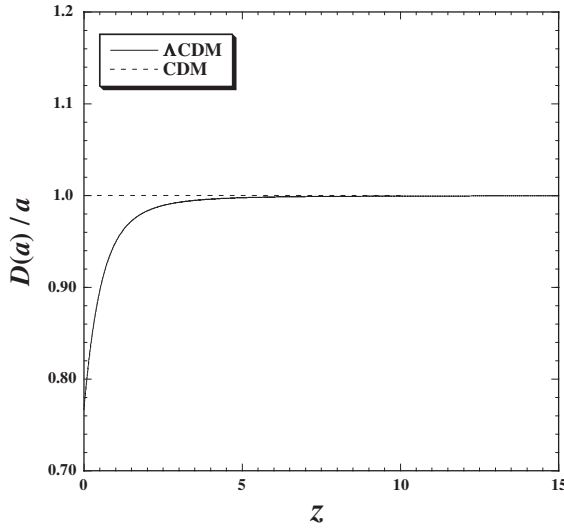


Figure 4.1 The evolution of the quantity $D(a)/a$ versus the redshift z for the Λ CDM model with $\Omega_m^{(0)} = 0.28$ and the CDM model with $\Omega_m^{(0)} = 1.0$. Since the gravitational potential Φ does not vary in time during the matter-dominated era, the quantity $D(a)/a$ remains constant for the CDM model. However, in the Λ CDM model, the decrease of the gravitational potential at late times leads to the variation of $D(a)/a$.

in which dark energy does not cluster, we can also ignore the dark energy perturbation compared to the matter perturbation. Under the sub-horizon approximation ($k \gg \mathcal{H}$), Eq. (4.181) then reduces to

$$k^2 \Phi = 4\pi G a^2 \rho_m \delta_m. \quad (4.210)$$

Using the relations $\rho_m = \rho_m^{(0)}/a^3$ and $\Omega_m^{(0)} = 8\pi G \rho_m^{(0)}/(3H_0^2)$, Eq. (4.210) can be expressed as

$$\delta_m(k, a) = \frac{2k^2 a}{3\Omega_m^{(0)} H_0^2} \Phi(k, a). \quad (4.211)$$

Using Eqs. (4.194), (4.209), and (4.211), the power spectrum of matter perturbations at the present epoch is

$$P_{\delta_m} \equiv \langle |\delta_m(k, a_0)|^2 \rangle = \frac{2\pi^2 \delta_H^2}{(\Omega_m^{(0)})^2} \left(\frac{k}{H_0} \right)^{n_s} T^2(k) D^2(a_0) H_0^{-3}. \quad (4.212)$$

On large scales characterized by $x = k/k_{\text{eq}} \ll 1$ the matter power spectrum has a scale-dependence $P_{\delta_m} \propto k^{n_s}$, which grows for increasing k . On small scales ($x = k/k_{\text{eq}} \gg 1$) one has $P_{\delta_m} \propto k^{n_s-4} (\ln k)^2$, which decreases for increasing k . Therefore there is a peak for the matter power spectrum, whose wavenumber is

determined by k_{eq} given in Eq. (4.195). As we will see in Section 5.5, the matter power spectrum is important in measuring dark energy properties.

4.11 Perturbed photon propagation

So far we have studied the evolution of matter density perturbations. However, what we really observe in cosmology is radiation emitted from sources. It is then time to face the question of the propagation of radiation in a perturbed Universe. That is, what happens to photons (their energy, their path) when they propagate through a perturbed space time?

The general answer is provided by the two equations of light propagation (we do not consider polarization here). If $k^\mu = dx^\mu/d\lambda_s$ is the photon momentum, then the null condition and the geodesic equation read

$$k^\mu k_\mu = 0, \quad (4.213)$$

$$\frac{dk^\mu}{d\lambda_s} + \Gamma_{\alpha\beta}^\mu k^\alpha k^\beta = 0, \quad (4.214)$$

where λ_s is an affine parameter that plays no explicit role because we can always use the $\mu = 0$ geodesic equation to convert λ_s to the conformal time η . Solving these equations in the perturbed metric (4.13) gives the general equations of photon propagation. The solution will give the variation in the photon's frequency and path due to the inhomogeneous metric. By observing the inhomogeneities in frequency and in the angular direction, we can constrain the perturbations that the photon went through during its long cosmic ride. We will see in Chapter 14 that these effects, in particular the ISW effect and the weak lensing, are very useful cosmological probes.

Let us start by splitting the momentum vector $k^\mu = dx^\mu/d\lambda_s$ into a background and a perturbed value

$$k^\mu = \hat{k}^\mu + \delta k^\mu. \quad (4.215)$$

We consider a photon that, in the unperturbed metric, propagates along the direction r , so that the background propagation equation in flat space is $d\eta = dr$. In this section the perturbed quantities are in real space, not in Fourier space, since we need to deal with physical trajectories. Also we use the explicit derivation symbol $\partial/\partial\eta$ instead of the prime because we want to distinguish carefully between partial and total derivatives.

At background level, the geodesic equation (index 0) gives simply

$$\frac{d\hat{k}^0}{d\lambda_s} = -2\mathcal{H}(\hat{k}^0)^2, \quad (4.216)$$

where we have used $\Gamma_{00}^0 = \mathcal{H}$ for the flat FLRW metric with the conformal time η . Equation (4.216) is integrated to give

$$\hat{k}^0 = \frac{d\eta}{d\lambda_s} \propto a^{-2}. \quad (4.217)$$

This is all we need to convert λ_s to η . Since the photon frequency is $\nu \equiv dt/d\lambda_s = ad\eta/d\lambda_s$, we see that the frequency redshifts in proportion to $1/a$, as it should. We also need the perturbed null-geodesic condition

$$\delta k^\mu \hat{k}_\mu + \hat{k}^\mu \delta k_\mu = 0. \quad (4.218)$$

For radial trajectories the null path condition $ds^2 = 0$ gives

$$(1 + \Psi)d\eta = (1 + \Phi)dr, \quad (4.219)$$

or

$$\frac{d\eta}{d\lambda_s} = (1 + \Phi - \Psi) \frac{dr}{d\lambda_s}, \quad (4.220)$$

which gives

$$\hat{k}^0 + \delta k^0 = (1 + \Phi - \Psi)(\hat{k}^r + \delta k^r). \quad (4.221)$$

We obtain $\hat{k}^0 = \hat{k}^r$ from the zero-th order equation. The first-order terms give

$$\delta k^0 = \delta k^r + \hat{k}^0(\Phi - \Psi). \quad (4.222)$$

Now we may be tempted to derive the first-order perturbed $\mu = 0$ equation (4.214) straightforwardly as $d\delta k^0/d\lambda_s = -4\mathcal{H}\hat{k}^0\delta k^0 - (\hat{k}^0)^2(\partial\Phi/\partial\eta + \partial\Psi/\partial\eta + 2\Psi_{,r})$. However, that would be wrong, since the affine parameter λ_s needs to be perturbed as well. We cannot take it to be the same as for the zeroth-order case. The way to proceed is then to derive the full expression up to first-order:

$$\frac{dk^0}{d\lambda_s} = -2\mathcal{H}[(\hat{k}^0)^2 + 2\hat{k}^0\delta k^0] - (\hat{k}^0)^2 \left(\frac{\partial\Phi}{\partial\eta} + \frac{\partial\Psi}{\partial\eta} + 2\Psi_{,r} \right), \quad (4.223)$$

where we used Eq. (4.222) and where $k^0 = \hat{k}^0 + \delta k^0$, so there is only one affine parameter for the full momentum k^0 . The situation is simplified if we “complete” the expression $(\hat{k}^0)^2 + 2\hat{k}^0\delta k^0$ with an extra second-order term $(\delta k^0)^2$ to obtain $(k^0)^2$ and similarly substitute $(k^0)^2$ for $(\hat{k}^0)^2$. This is allowed because we are adding second-order terms that will be automatically discarded later on. Then we obtain

$$\frac{dk^0}{d\lambda_s} = -2\mathcal{H}(k^0)^2 - (k^0)^2 \left(\frac{\partial\Phi}{\partial\eta} + \frac{\partial\Psi}{\partial\eta} + 2\Psi_{,r} \right). \quad (4.224)$$

Dividing Eq. (4.224) with respect to $k^0 = d\eta/d\lambda_s$, we can rewrite it in the form

$$\frac{1}{a^2 k^0} \frac{d(a^2 k^0)}{d\eta} = - \left(\frac{\partial \Phi}{\partial \eta} + \frac{\partial \Psi}{\partial \eta} + 2\Psi_{,r} \right). \quad (4.225)$$

Now we split the momentum k^0 to be $k^0 = \hat{k}^0 + \delta k^0$. Using the fact that Eq. (4.216) is equivalent to $d(a^2 \hat{k}^0)/d\eta = 0$, it follows that

$$\frac{1}{a^2 k^0} \frac{d(a^2 k^0)}{d\eta} \rightarrow \frac{d(\delta k^0 / k^0)}{d\eta}, \quad (4.226)$$

where, at first-order, we can put equivalently k^0 or $\hat{k}^0$ at the denominator. We then obtain the following equation

$$\frac{d(\delta k^0 / k^0)}{d\eta} = - \left(\frac{\partial \Phi}{\partial \eta} + \frac{\partial \Psi}{\partial \eta} + 2\Psi_{,r} \right). \quad (4.227)$$

For the spatial equations we need to derive only the two directions x^1 and x^2 orthogonal to the propagation direction r . For $i = 1, 2$, using the geodesic condition $d\eta^2 - dr^2 = 0$, we find from Eq. (4.214)

$$\frac{d^2 x^i}{d\lambda_s^2} + 2\mathcal{H} \frac{d\eta}{d\lambda_s} \frac{dx^i}{d\lambda_s} = \left(\frac{d\eta}{d\lambda_s} \right)^2 (\Phi_{,i} - \Psi_{,i}). \quad (4.228)$$

Here we do not need to worry about the unperturbed λ_s because there is no unperturbed propagation along the directions x^1 and x^2 . We write this equation explicitly in terms of the perturbed coordinates $x^i(\lambda_s)$ of the light ray instead of k^i since this will be useful in the weak lensing observation. Notice that in several derivations at this stage one encounters additional terms proportional to $k^i \Phi_{,i}$ which are then eliminated by invoking the so-called thin-lens approximation. Since we are considering first-order quantities, these terms are automatically canceled from the beginning.

Now the problem splits neatly in two: (i) the solution to Eq. (4.227) leads to the discussion of the Sachs–Wolfe effect, i.e. the change of a photon's redshift in passing through a gravitational potential, (ii) the solution to Eq. (4.228) leads to weak lensing, i.e. the deviation of a light ray passing through the same.

4.11.1 The Sachs–Wolfe effect

Let us begin with Eq. (4.227). First we notice the following relation

$$\hat{k}^0 \left(\frac{\partial \Psi}{\partial \eta} + \Psi_{,r} \right) = \frac{\partial \Psi}{\partial \eta} \frac{d\eta}{d\lambda_s} + \frac{\partial \Psi}{\partial r} \frac{dr}{d\lambda_s} = \frac{d\Psi}{d\lambda_s} \rightarrow \frac{\partial \Psi}{\partial \eta} + \Psi_{,r} = \frac{d\Psi}{d\eta}, \quad (4.229)$$

where we have used $\hat{k}^0 = d\eta/d\lambda_s$ and also $d\eta = dr$ along the unperturbed light ray. We then find that Eq. (4.227) can be written as

$$\frac{d(\delta k^0/k^0)}{d\eta} = -2\frac{d\Psi}{d\eta} - \left(\frac{\partial\Phi}{\partial\eta} - \frac{\partial\Psi}{\partial\eta} \right), \quad (4.230)$$

which is integrated to give

$$\frac{\delta k^0}{k^0} \Big|_E^O = -2\Psi|_E^O - \int_E^O \left(\frac{\partial\Phi}{\partial\eta} - \frac{\partial\Psi}{\partial\eta} \right) d\eta, \quad (4.231)$$

where “ O ” and “ E ” represent the instants of observation and emission, respectively, and $\Psi|_E^O \equiv \Psi_O - \Psi_E$. Notice that the last term does not integrate to $(\Phi - \Psi)_E^O$ since the potentials must be calculated along the light-ray null path $r(\eta)$ and the integral contains the partial, not total, derivative with respect to η .

This is our first result. The frequency shifts by an amount that depends on the difference between the potential at emission and at observation and on the line-of-sight integral of $\Phi - \Psi$. The first effect is referred to as the *Sachs–Wolfe effect* (although often another term gets counted as Sachs–Wolfe, see below) and the second as the *integrated Sachs–Wolfe effect*. The most important application of this result is the calculation of the CMB temperature anisotropies.

The temperature of a black body distribution of photons is proportional to their average frequency $\hat{\nu}$. More exactly, for photons propagating with four-momentum k^μ and emitted by a body moving with respect to us with velocity u_μ , the temperature we measure is proportional to $-k^\mu u_\mu = \hat{\nu}$. The same applies if u_μ is our own velocity with respect to the source. Therefore if a photon is emitted at E and observed at O , the emission temperature T_E and observed temperature T_O are related with each other via

$$\frac{T_O}{T_E} = \frac{(k^\mu u_\mu)_O}{(k^\mu u_\mu)_E}. \quad (4.232)$$

The observed temperature fluctuation is

$$\frac{\delta T}{T} \Big|_O = \frac{\delta T}{T} \Big|_E + \frac{\delta(k^\mu u_\mu)}{k^\mu u_\mu} \Big|_O - \frac{\delta(k^\mu u_\mu)}{k^\mu u_\mu} \Big|_E. \quad (4.233)$$

As usual we assume a radial propagation along the coordinate r , i.e. $d\eta = dr$. Recall that the four-velocity is given by $u_\mu = [-a(1 + \Psi), av_i]$ for the perturbed metric (4.13). Notice that at zero-th order we have $k^\mu u_\mu = k^0 u_0$. Let us consider

the second term on the r.h.s. of Eq. (4.233). The non-zero first-order elements are

$$\frac{k^0 \delta u_0}{k^\mu u_\mu} = \frac{\delta u_0}{u_0} = \Psi, \quad (4.234)$$

$$\frac{k^i \delta u_i}{k^\mu u_\mu} = \frac{k^i \delta u_i}{k^0 u_0} = -\frac{k^i}{k^0} v_i = e^i v_i, \quad (4.235)$$

where $e^i \equiv -k^i/k^0$ is the unit direction vector, and the term $(\delta k^0)u_0/(k^\mu u_\mu) = \delta k^0/k^0$ is given in Eq. (4.231).

We can now put the pieces together. We find

$$\left. \frac{\delta T}{T} \right|_O = \left. \frac{\delta T}{T} \right|_E + e^i v_i|_E^O - \Psi|_E^O + \int_E^O \left(\frac{\partial \Psi}{\partial \eta} - \frac{\partial \Phi}{\partial \eta} \right) d\eta. \quad (4.236)$$

The term $\Psi|_E^O = \Psi_O - \Psi_E$ can be further simplified since the temperature distortion induced by the local gravitational potential Ψ_O is unobservable (it affects equally the whole CMB). Then we are left with four other terms in Eq. (4.236), which in turn are called the intrinsic temperature fluctuation ($\delta T/T|_E$), the Doppler shift ($e^i v_i|_E^O$), the Sachs–Wolfe (SW) effect (Ψ_E), and the integrated Sachs–Wolfe (ISW) effect ($\int_E^O (\partial \Psi/\partial \eta - \partial \Phi/\partial \eta) d\eta$), respectively. When we discuss photons coming from the last scattering surface of the CMB, the intrinsic temperature fluctuation is also counted as the SW effect (the separation between these effects is in fact gauge dependent). As we have mentioned in Section 4.10 the adiabatic initial condition of the baryon–photon fluid system in the tight coupling regime corresponds to $\delta_m = 3\delta T/T|_E$, where δ_m is the matter perturbation. From Eq. (4.71) the matter perturbation satisfies the relation $\delta_m = -2\Psi$ on scales larger than the Hubble radius as long as the matter dominates the overall fluctuations and the anisotropic stress is neglected such that $\Phi = -\Psi$. So finally one arrives at the (adiabatic, large-scale, shear-free) familiar result

$$\left. \frac{\delta T}{T} \right|_{\text{SW}} = -\frac{2}{3}\Psi_E + \Psi_E = \frac{1}{3}\Psi_E, \quad (4.237)$$

where the subscript E refers to the epoch of last scattering. Since the dark energy density in most models is negligible at early times, Ψ_E is hardly affected by it.

In contrast, the last term in Eq. (4.236), the ISW term, is a line-of-sight term and does therefore contain potentially much information on the recent Universe. Moreover, we have already seen that the gravitational potential is constant for a matter-dominated Universe and therefore yields no ISW signal. This shows that the ISW effect is a direct diagnostic of something which is not ordinary pressureless matter, just as we are looking for. The constraint on dark energy models through the ISW effect will be further discussed in Section 14.1.1.

4.11.2 Weak lensing

Let us now tackle the spatial part of the geodesic equations. Using Eqs. (4.228) and (4.217), we obtain the propagation equation for $i = 1, 2$:

$$\frac{d^2 x^i}{dr^2} = \psi_{,i} , \quad (4.238)$$

where ψ is the lensing potential defined by

$$\psi \equiv \Phi - \Psi . \quad (4.239)$$

In standard General Relativity with ordinary matter, $\psi = -2\Psi$. Since the displacement vector $\mathbf{x} = (x^1, x^2)$ is small, we can put $x^i = r\theta^i$ and hence Eq. (4.238) is written as

$$\frac{d^2}{dr^2}(r\theta^i) = \psi_{,i} . \quad (4.240)$$

If the light ray reaches the observer located at $r = 0$ through the direction $\theta_0^i = (\theta_0^1, \theta_0^2)$ the integration of Eq. (4.240) twice leads to (here and in the rest of this section we use primes to denote dummy integration variables, not derivatives)

$$\theta^i = \theta_0^i + \frac{1}{r} \int_0^r dr'' \int_0^{r'} dr' \psi_{,i}(r'\theta_0^1, r'\theta_0^2, r') , \quad (4.241)$$

where the integration constant is to be equal to the observed angle θ_0^i so that the angle remains the same in the absence of metric perturbations ($\psi = 0$). The region of the integral with respect to r' and r'' is restricted to be in the region $0 < r'' < r, 0 < r' < r''$, or in other words, $r' < r'' < r, 0 < r' < r$. Carrying out the r'' integral of Eq. (4.241) in the latter region, it follows that

$$\theta^i = \theta_0^i + \int_0^r dr' \left(1 - \frac{r'}{r}\right) \psi_{,i}(r'\theta_0^1, r'\theta_0^2, r') . \quad (4.242)$$

Two light rays separated by a small interval Δx will obey the equation

$$\Delta\theta^i = \Delta\theta_0^i + \Delta\theta_0^j \int_0^r dr' \left(1 - \frac{r'}{r}\right) r' \psi_{,ij}(r'\theta_0^1, r'\theta_0^2, r') , \quad (4.243)$$

where the term $r' \psi_{,ij}$ arises by taking the variation of $\psi_{,i}$ with respect to θ_0^j ($j = 1, 2$). Therefore, if the separation $\Delta\theta^i$ is taken on the source plane at $r = r_s$, we have an equation that connects the source plane with the observation plane at $r = 0$. The entire phenomenon can be described by the symmetric transformation matrix (see Fig. 4.2)

$$A_{ij} \equiv \frac{\partial \theta_s^i}{\partial \theta_0^j} = \delta_{ij} + D_{ij} , \quad (4.244)$$

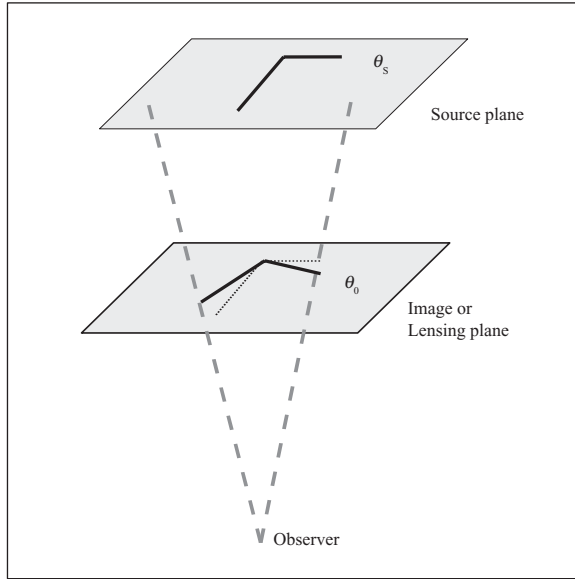


Figure 4.2 The geometry of the weak lensing.

where

$$D_{ij} = \int_0^{r_s} dr' \left(1 - \frac{r'}{r_s} \right) r' \psi_{,ij} = \begin{pmatrix} -\kappa_{wl} - \gamma_1 & -\gamma_2 \\ -\gamma_2 & -\kappa_{wl} + \gamma_1 \end{pmatrix}, \quad (4.245)$$

is the distortion tensor. The parameter

$$\kappa_{wl} = -\frac{1}{2} \int_0^{r_s} dr' \left(1 - \frac{r'}{r_s} \right) r' (\psi_{,11} + \psi_{,22}), \quad (4.246)$$

is called the *convergence* and describes the magnification of the source image. The quantities γ_1, γ_2 are the two components of the *shear field*

$$\gamma_1 = -\frac{1}{2} \int_0^{r_s} dr' \left(1 - \frac{r'}{r_s} \right) r' (\psi_{,11} - \psi_{,22}), \quad (4.247)$$

$$\gamma_2 = -\int_0^{r_s} dr' \left(1 - \frac{r'}{r_s} \right) r' \psi_{,12}, \quad (4.248)$$

and describe the distortion of the source image. Although in principle both magnification and shape distortion could be used as cosmological tools, the noise in the former (i.e. the large intrinsic variation of galaxy luminosities) has not allowed so far a practical use of the magnification in cosmology. In Section 14.4 we will study how to employ the distortion as a test of cosmology.

4.12 Problems

4.1 Let the distribution function of the density contrast δ_x be given by the Gaussian

$$P_x(\delta_x) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{\delta_x^2}{2\sigma^2}\right). \quad (4.249)$$

Suppose that the fluctuations δ_x, δ_y at any two given points x, y separated by r are distributed as a two-dimensional Gaussian field with the probability distribution

$$P_{xy}(\delta_x, \delta_y) = \frac{1}{2\pi(\sigma^4 - \xi_{12}^2)^{1/2}} \exp\left[-\frac{1}{2}\delta_i C_{ij}^{-1} \delta_j\right], \quad (4.250)$$

where $\delta_i = (\delta_x, \delta_y)$ and the covariance matrix is

$$C_{ij} \equiv \int \delta_i \delta_j P_{xy}(\delta_x, \delta_y) d\delta_x d\delta_y = \begin{pmatrix} \sigma^2 & \xi_{12}(r) \\ \xi_{12}(r) & \sigma^2 \end{pmatrix}. \quad (4.251)$$

Let us now cut the field so that we consider only fluctuations above a given threshold.

- (1) Find the general expression of the probability of having $\delta_x > \nu\sigma$ given that $\delta_y > \nu\sigma$ at a distance $|x - y| = r$, where ν is a positive parameter that defines a threshold in units of σ .
- (2) Suppose that we divide a galaxy distribution into many small equal-volume regions above threshold ($\delta > \nu\sigma$) and below threshold. Find the expected value of the fraction of regions above threshold within a distance r of a region above threshold, divide this for the fraction of regions above threshold and interpret this as the correlation function $1 + \xi_{>b}$ of regions above threshold.
- (3) Find the limit of $\xi_{12} \ll 1$ and $\nu \gg 1$ and interpret $b = \nu/\sigma$ as the bias of regions above threshold with respect to the general field. (From Ref. [101].)

Useful relations are:

$$\frac{1}{2}\text{erfc}\left(\frac{\nu}{\sqrt{2}}\right) \equiv \frac{1}{\sqrt{2\pi}\sigma} \int_{\nu\sigma}^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx, \quad (4.252)$$

$$\text{erfc}(x) \simeq \frac{e^{-x^2}}{x\sqrt{\pi}}, \quad \text{for } x \gg 1, \quad (4.253)$$

$$\text{erfc}[x(1 + \varepsilon)] \simeq \text{erfc}(x) - \frac{2}{\sqrt{\pi}} \varepsilon x e^{-x^2}, \quad \text{for } \varepsilon \ll 1. \quad (4.254)$$

4.2 Derive Eqs. (4.176) and (4.177) from Eq. (4.170) by neglecting the contribution of the quadrupole moment Θ_2 .

4.3 Confirm that Eq. (4.179) follows from Eqs. (4.176)–(4.178).

5

Observational evidence of dark energy

The existence of dark energy is supported by a number of observations. This includes (i) the age of the Universe compared to oldest stars, (ii) supernovae observations, (iii) Cosmic Microwave Background (CMB), (iv) baryon acoustic oscillations (BAO), and (v) large-scale structure (LSS).

Even before 1998 it was known that in a CDM Universe the cosmic age can be smaller than the age of the oldest stars. Dark energy can account for this discrepancy because its presence can make the cosmic age longer. The first strong evidence for the acceleration of the Universe today came however by measuring the luminosity distance of the type Ia supernovae (SN Ia). The CMB observations are also consistent with the presence of dark energy, although the constraint coming from the CMB alone is not so strong. The measurements of BAO have provided another independent test for the existence of dark energy. The power spectrum of matter distributions also favors a Universe with dark energy rather than the CDM Universe. In the following we shall discuss this observational evidence for dark energy. The statistical method used to constrain cosmological parameters will be discussed in Chapter 13. More details on present and future observational aspects to detect dark energy will be presented in Chapter 14.

5.1 The age of the Universe

As we already mentioned, the inverse of the Hubble constant H_0 is a rough measure of the age t_0 of the Universe. Here we shall compute t_0 more precisely and compare it with the age of the oldest stars. For simplicity we assume that the equation of state of dark energy is a constant, in which case we have $\rho_{\text{DE}} = \rho_{\text{DE}}^{(0)}(1+z)^{3(1+w_{\text{DE}})}$ from Eq. (2.81). Taking into account radiation, non-relativistic matter, and dark energy as components of the Universe, Eq. (2.84) gives the Hubble parameter $H(z)$

normalized by H_0 :

$$E(z) = \left[\Omega_r^{(0)}(1+z)^4 + \Omega_m^{(0)}(1+z)^3 + \Omega_{\text{DE}}^{(0)}(1+z)^{3(1+w_{\text{DE}})} + \Omega_K^{(0)}(1+z)^2 \right]^{1/2}. \quad (5.1)$$

On using the relation $dt = -dz/[(1+z)H]$, the age of the Universe is expressed as

$$t_0 = H_0^{-1} \int_0^\infty \frac{dz}{E(z)(1+z)}. \quad (5.2)$$

The integral (5.2) is dominated by the terms at low redshifts. Since $\Omega_r^{(0)}$ is of the order of 10^{-5} – 10^{-4} from Eq. (2.47), radiation becomes important only for high redshifts ($z \gtrsim 1000$). Hence it is a good approximation to neglect the contribution from radiation when we evaluate Eq. (5.2). Let us consider the case of the cosmological constant ($w_{\text{DE}} = -1$). Then the age of the Universe is given by

$$t_0 = H_0^{-1} \int_1^\infty \frac{dx}{x \left[\Omega_m^{(0)} x^3 + \Omega_{\text{DE}}^{(0)} + \Omega_K^{(0)} x^2 \right]^{1/2}}, \quad (5.3)$$

where $x \equiv 1+z$ and $\Omega_m^{(0)} + \Omega_{\text{DE}}^{(0)} + \Omega_K^{(0)} = 1$.

For the flat Universe ($\Omega_K^{(0)} = 0$), Eq. (5.3) is integrated to give

$$t_0 = \frac{H_0^{-1}}{3\sqrt{1-\Omega_m^{(0)}}} \ln \left(\frac{1 + \sqrt{1-\Omega_m^{(0)}}}{1 - \sqrt{1-\Omega_m^{(0)}}} \right), \quad (5.4)$$

where we have used the relation $\Omega_m^{(0)} + \Omega_{\text{DE}}^{(0)} = 1$. In the limit $\Omega_{\text{DE}}^{(0)} \rightarrow 0$ we have

$$t_0 = \frac{2}{3} H_0^{-1}. \quad (5.5)$$

On using the value (2.36) together with $h = 0.72 \pm 0.08$, the age of the Universe in the absence of the cosmological constant is in the range $8.2 \text{ Gyr} < t_0 < 10.2 \text{ Gyr}$. Carretta *et al.* [102] estimated the age of globular clusters in the Milky Way to be $12.9 \pm 2.9 \text{ Gyr}$, whereas Jimenez *et al.* [103] obtained the value $13.5 \pm 2 \text{ Gyr}$. Hansen *et al.* [104] constrained the age of the globular cluster M4 to be $12.7 \pm 0.7 \text{ Gyr}$ by using the method of the white dwarf cooling sequence. In most cases the ages of globular clusters are larger than 11 Gyr. Then the cosmic age estimated by Eq. (5.5) is inconsistent with the ages of the oldest globular clusters.

This problem can be circumvented by taking into account the cosmological constant (or dark energy with an equation of state w_{DE} close to -1). Equation (5.4) shows that t_0 gets larger for decreasing $\Omega_m^{(0)}$. In the limit $\Omega_m^{(0)} \rightarrow 0$ we have $t_0 \rightarrow \infty$. In Fig. 5.1 we plot the cosmic age (5.4) versus $\Omega_m^{(0)}$ together with the

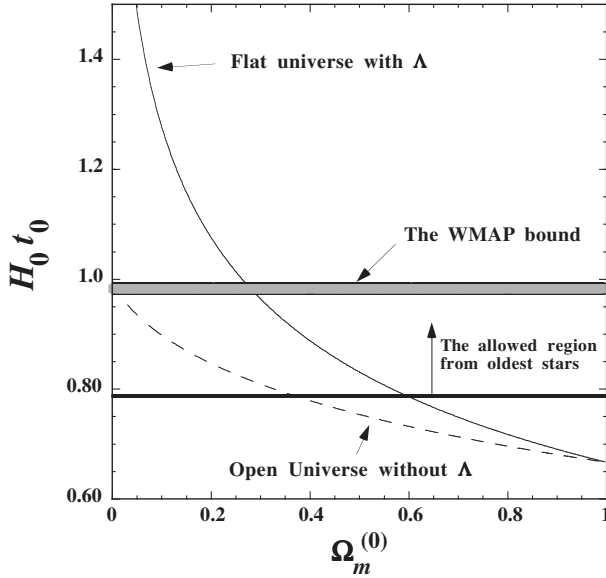


Figure 5.1 The cosmic age t_0 in the unit of H_0^{-1} versus $\Omega_m^{(0)}$. The thin-solid curve describes a flat Universe in the presence of the cosmological constant Λ with the relation $\Omega_m^{(0)} + \Omega_{\text{DE}}^{(0)} = 1$. The dashed curve corresponds to an open Universe without the cosmological constant. The thick-solid line is a minimum age allowed from the ages of oldest globular clusters (> 11 Gyr). We also show the bound coming from the WMAP 5-year data with $h = 0.70$. The flat Universe with the cosmological constant is consistent with the WMAP bound for $0.271 < \Omega_m^{(0)} < 0.289$.

boundary allowed from the oldest stellar ages. In order to satisfy the condition $t_0 > 11$ Gyr we require that $0 < \Omega_m^{(0)} < 0.55$. The WMAP 5-year constraint on the cosmic age (assuming the Λ CDM model) is given by $t_0 = 13.73 \pm 0.12$ Gyr [15]. Under this bound we find that the density parameter of non-relativistic matter is constrained to be $0.245 < \Omega_m^{(0)} < 0.261$ for $h = 0.72$. Of course this bound changes if different values of h are chosen. When $h = 0.70$ the constraint becomes $0.271 < \Omega_m^{(0)} < 0.289$ (see Fig. 5.1).

In the open Universe ($\Omega_K^{(0)} > 0$) it is also possible to make the cosmic age larger than $(2/3)H_0^{-1}$ even in the absence of dark energy. Setting $\Omega_{\text{DE}}^{(0)} = 0$ in Eq. (5.3), we obtain

$$t_0 = \frac{H_0^{-1}}{1 - \Omega_m^{(0)}} \left[1 + \frac{\Omega_m^{(0)}}{2\sqrt{1 - \Omega_m^{(0)}}} \ln \left(\frac{1 - \sqrt{1 - \Omega_m^{(0)}}}{1 + \sqrt{1 - \Omega_m^{(0)}}} \right) \right], \quad (5.6)$$

where $\Omega_m^{(0)} + \Omega_K^{(0)} = 1$. In the limit $\Omega_m^{(0)} \rightarrow 1$ we recover the value (5.5) in the flat Universe. Meanwhile, in the limit $\Omega_m^{(0)} \rightarrow 0$, we have $t_0 \rightarrow H_0^{-1}$. The cosmic

age in the open Universe does not become so large compared to the case of the flat Universe with the cosmological constant. Since the curvature $|\Omega_K^{(0)}|$ has been constrained to be much smaller than unity from the WMAP measurements [15], it is not possible to satisfy the condition $t_0 > 11$ Gyr for $h = 0.72 \pm 0.08$ in the open Universe without dark energy.

The above discussion shows that the existence of dark energy is crucially important to solve the cosmic age problem.

5.2 Supernova observations

In 1998 Riess *et al.* [High-redshift Supernova Search Team (HSST)] [1] and Perlmutter *et al.* [Supernova Cosmology Project (SCP)] [2] independently reported the late-time cosmic acceleration by observing distant supernovae of type Ia (SN Ia). Up to 1998 Riess *et al.* had discovered 16 high-redshift SN Ia together with 34 nearby supernovae, while Perlmutter *et al.* had found 42 supernovae in the redshift range $z = 0.18\text{--}0.83$.

The explosion of supernovae is extremely luminous and causes a burst of radiation. The supernovae can be classified according to the absorption lines of chemical elements. If the spectrum of a supernova includes a spectral line of hydrogen, it is classified Type II. Otherwise it is called Type I. If a supernova contains an absorption line of singly ionized silicon, it is further classified Type Ia (note that Type Ib contains a line of helium, whereas Type Ic lacks the lines of both silicon and helium). The explosion of Type Ia occurs when the mass of a white dwarf in a binary system exceeds the Chandrasekhar limit [105] by absorbing gas from another star. Since the absolute luminosity of Type Ia is almost constant at the peak of brightness, the distance to a SN Ia can be determined by measuring its observed (apparent) luminosity. Thus the SN Ia is a kind of “standard candle” by which luminosity distance can be measured observationally.

In reality things are more complicated than this simple view. The intrinsic spread in absolute magnitudes is actually too large to produce stringent cosmological constraints. However, at the end of the 1990s, a high-quality sample of “local” (i.e. $z \ll 1$) supernovae allowed the absolute magnitude to be correlated with the width of the light curve [106]: brighter supernovae have a broader light curve. By measuring at the same time the apparent magnitude and the light curve it is possible therefore to predict the absolute magnitude. Although in the following we refer to a universal SN Ia absolute magnitude, we always mean the magnitude corrected for the light curve width.

The apparent magnitude, m , is often used as a measure of brightness of stars observed on Earth. Let us consider two stellar objects whose apparent fluxes are given by $\mathcal{F}_1$ and $\mathcal{F}_2$. The apparent magnitudes of those stars (m_1 and m_2) are related

to the fluxes according to

$$m_1 - m_2 = -\frac{5}{2} \log_{10} \left(\frac{\mathcal{F}_1}{\mathcal{F}_2} \right). \quad (5.7)$$

This implies that a star with $m_1 = 1$ is about 100 times brighter than one with $m_2 = 6$. From the definition (5.7) the apparent magnitude is smaller for brighter objects. Choosing an appropriate reference flux, for Sun and Moon we have $m = -26.7$ and $m = -12.6$, respectively.

We define the absolute magnitude M of an object in terms of an apparent magnitude m and a luminosity distance d_L :

$$m - M = 5 \log_{10} \left(\frac{d_L}{10 \text{ pc}} \right). \quad (5.8)$$

If the distance is expressed in Megaparsec then the relation can be written as

$$m - M = 5 \log_{10} d_L + 25. \quad (5.9)$$

In other words the absolute magnitude corresponds to the apparent magnitude the object would have if it were located at the luminosity distance $d_L = 10 \text{ pc}$ from the observer. An additional correction, denoted as K -correction, is due to the fact that as the redshift increases we observe different parts of the source spectrum: we always assume that this correction has been already included in the estimation of m .

The absolute magnitude of SN Ia is known to be around $M = -19$ at the peak of brightness. If we consider two SN Ia whose apparent magnitudes are m_1 and m_2 with luminosity distances d_{L_1} and d_{L_2} , respectively, we obtain the following relation from Eq. (5.9):

$$m_1 - m_2 = 5 \log_{10} \left(\frac{d_{L_1}}{d_{L_2}} \right). \quad (5.10)$$

Since the observed flux $\mathcal{F}$ is proportional to d_L^{-2} from Eq. (2.63), we find that Eq. (5.7) is consistent with Eq. (5.10).

Since the (corrected) peak absolute magnitude M is the same for any SN Ia under the assumption of standard candles, the luminosity distance $d_L(z)$ is obtained from Eq. (5.9) by observing the apparent magnitude m . The redshift z of the corresponding SN Ia can be found by measuring the wavelength λ of light [see Eq. (2.28)]. The observations of many SN Ia provide the dependence of the luminosity distance d_L in terms of z . Comparing observational data with the theoretical distance (2.68), it is possible to know the expansion history of the Universe for the redshift $z \lesssim \mathcal{O}(1)$.

Let us consider the case in which the Universe is dominated by a non-relativistic fluid and dark energy with an equation of state w_{DE} . In this case the Hubble parameter $H(z)$ is given by Eq. (2.84) with $\Omega_r^{(0)} \simeq 0$. Using the expansion (2.69)

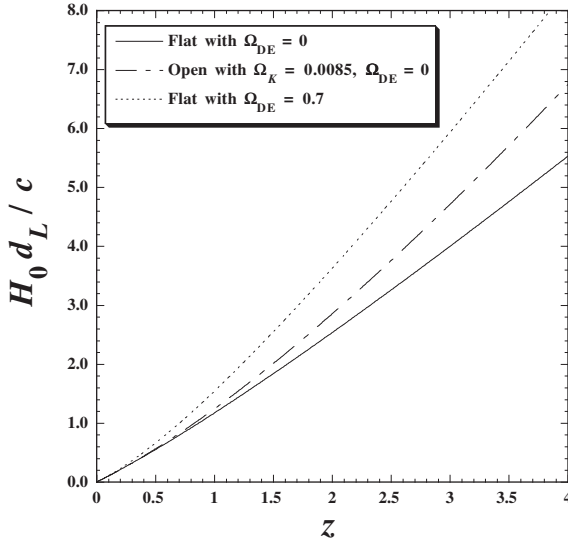


Figure 5.2 The luminosity distance d_L versus the redshift z for three cases: (a) a flat Universe without dark energy, (b) an open Universe ($\Omega_K^{(0)} = 0.0085$) without dark energy, and (c) a flat Universe with the cosmological constant ($\Omega_{DE}^{(0)} = 0.7$ and $w_{DE} = -1$). The presence of dark energy leads to a larger luminosity distance relative to the case without it. In the open Universe the luminosity distance also gets larger than that in the flat Universe.

around $z = 0$ we find that the luminosity distance, in the region $z \ll 1$, is given by

$$d_L(z) = \frac{c}{H_0} \left[z + \frac{1}{4} \left(1 - 3w_{DE}\Omega_{DE}^{(0)} + \Omega_K^{(0)} \right) z^2 + \mathcal{O}(z^3) \right]. \quad (5.11)$$

In the flat Universe without dark energy we have $d_L(z) = (c/H_0)[z + z^2/4 + \mathcal{O}(z^3)]$. In the presence of dark energy ($w_{DE} < 0$ and $\Omega_{DE}^{(0)} > 0$) the luminosity distance gets larger (see Fig. 5.2). Especially for smaller (negative) w_{DE} and for larger $\Omega_{DE}^{(0)}$ this tendency becomes more significant. In an open Universe ($K < 0$) the effect of the cosmic curvature also leads to a larger luminosity distance compared to the flat Universe. Since the curvature of the Universe is constrained to be close to the flat one ($-0.0175 < \Omega_K^{(0)} = -K/(a_0^2 H_0^2) < 0.0085$ [15]) from WMAP 5-year data, it is difficult to give rise to a significant difference relative to the flat Universe without dark energy. This property can be seen in Fig. 5.2, which shows that the difference is small in the region $z < 1.5$.

In 1998 Riess *et al.* [1] and Perlmutter *et al.* [2] released observational data of the apparent luminosity of high-redshift Type Ia supernovae ($0.2 \lesssim z \lesssim 0.8$). The data of low-redshift regions ($z < 0.1$) reported previously was also used in their analysis.

Let us pick up a few examples of data to understand how the luminosity distance is known observationally. First, consider two data of the apparent magnitudes in the low-redshift region of SN Ia: (i) 1990O: $m = 16.26$ ($z = 0.03$) and (ii) 1992bg: $m = 16.66$ ($z = 0.036$). Since the luminosity distance in the region $z \ll 1$ is well approximated by $d_L \simeq cz/H_0$ from Eq. (5.11), the absolute magnitude M is known from Eq. (5.9). We take the value $h = 0.7$ for the Hubble constant given in Eq. (2.33). We then obtain $M = -19.29$ and $M = -19.28$ for 1990O and 1992bg, respectively. This shows that the absolute luminosity of SN Ia is nearly constant ($M \simeq -19$), as we already mentioned.

Let us next use the high-redshift data reported by Perlmutter *et al.* [2]. Consider the two SN Ia data of the apparent magnitudes: (a) 1997R: $m = 23.83$ ($z = 0.657$), (b) 1995ck: $m = 23.57$ ($z = 0.656$). Employing the value $M = -19.15$ for the absolute magnitude, we find from Eq. (5.9) that the luminosity distance is given by $H_0 d_L/c = 0.920$ for 1997R and $H_0 d_L/c = 0.817$ for 1995ck. Notice that the approximation, $d_L \simeq cz/H_0$, is no longer valid in the high-redshift regime. Let us consider a flat Universe with a dark energy equation of state $w_{\text{DE}} = -1$ (i.e. the cosmological constant). Since $E(z) = [\Omega_m^{(0)}(1+z)^3 + \Omega_{\text{DE}}^{(0)}]^{1/2}$ in this case, the luminosity distance (2.68) reads

$$d_L(z) = \frac{c(1+z)}{H_0} \int_0^z \frac{d\tilde{z}}{[(1 - \Omega_{\text{DE}}^{(0)})(1+\tilde{z})^3 + \Omega_{\text{DE}}^{(0)}]^{1/2}}, \quad (5.12)$$

which can be evaluated numerically for given $\Omega_{\text{DE}}^{(0)}$. In order to satisfy the observational data $(H_0/c)d_L(z = 0.657) = 0.920$ for 1997R, we require that $\Omega_{\text{DE}}^{(0)} = 0.70$. Similarly we get $\Omega_{\text{DE}}^{(0)} = 0.38$ from the 1995ck data. Both data indicate the existence of dark energy.

Since observational data are prone to statistical and systematic errors, a few data are not enough to conclude that the present Universe is accelerating. Using 42 high-redshift SN Ia at redshifts between 0.18 and 0.83 together with 18 low-redshift SN Ia data from the Calan/Tololo Supernova Survey [106], Perlmutter *et al.* [2] showed that the cosmological constant is present at the 99% confidence level. They also found that the open Universe without the cosmological constant does not fit the data well. From Eq. (5.9) the apparent luminosity m gets larger for increasing luminosity distance d_L . Figure 5.3 shows that the observational data in the high-redshift regime favor the luminosity distance larger than the one predicted by the CDM model ($\Omega_m^{(0)} = 1$ and $\Omega_{\Lambda}^{(0)} = 0$). From the likelihood analysis of the SN Ia data accumulated by the year 1998, Perlmutter *et al.* found that the density parameter of non-relativistic matter is constrained to be $\Omega_m^{(0)} = 0.28_{-0.08}^{+0.09}$ (1σ statistical) in the flat Universe with the cosmological constant [2].

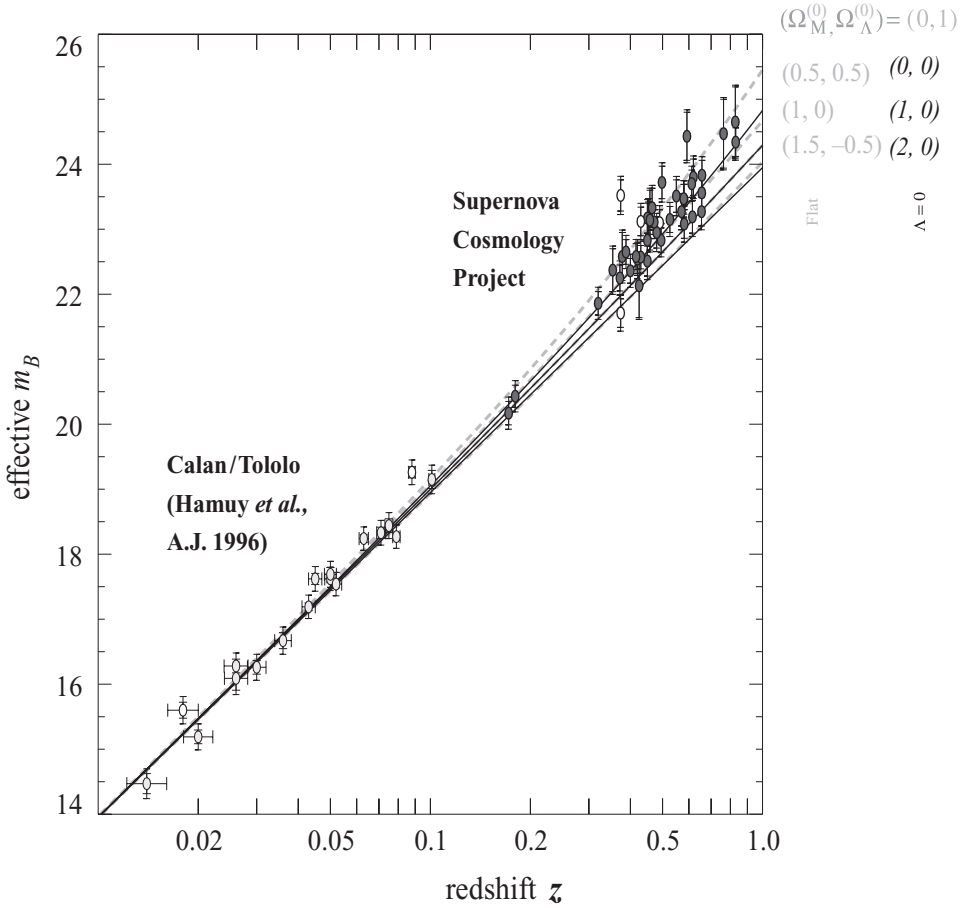


Figure 5.3 The effective apparent luminosity m_B versus the redshift z for 42 high-redshift SN Ia from the SCP [2] and 18 low-redshift SN Ia from the Calan/Tololo Supernova Survey [106]. Note that m_B involves the corrections of both sets for the SN Ia light-curve width-luminosity relation to the apparent luminosity m . The inner and outer error bars represent the uncertainty of measurements and the total uncertainty when the luminosity dispersion, 0.17 mag, of light-curve-width-corrected SN Ia is added in quadrature, respectively. The horizontal error bars show the assigned peculiar velocity uncertainty of 300 km s^{-1} . The solid curves are the theoretical prediction for m_B for a number of cosmological models without the cosmological constant: $(\Omega_m^{(0)}, \Omega_\Lambda^{(0)}) = (0, 0)$ (top), $(1, 0)$ (middle), and $(2, 0)$ (bottom). The dashed curves correspond to a number of flat cosmological models: $(\Omega_m^{(0)}, \Omega_\Lambda^{(0)}) = (0, 1)$ (top), $(0.5, 0.5)$ (second from top), $(1, 0)$ (third from top), and $(1.5, -0.5)$ (bottom). From Ref. [2].

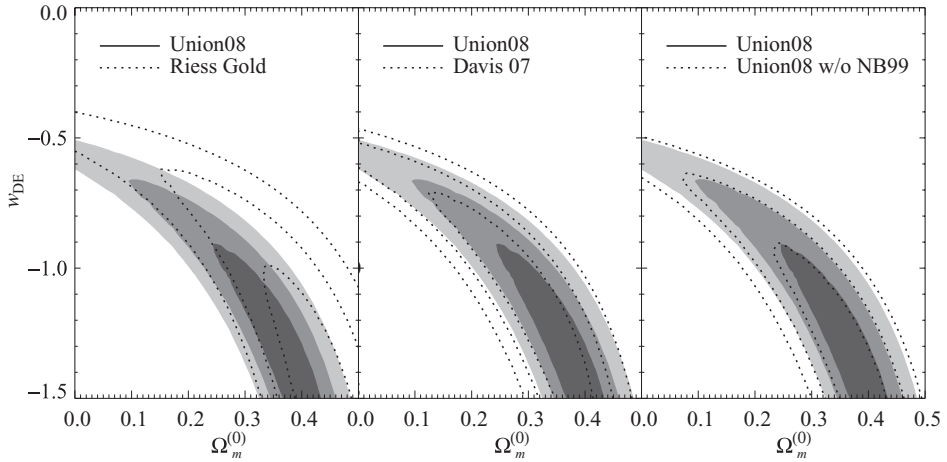


Figure 5.4 68.3%, 95.4%, and 99.7% confidence level contours on $(\Omega_m^{(0)}, w_{\text{DE}})$ from the SN Ia observations only for constant w_{DE} . In each column the filled contours correspond to the results coming from the Union data by Kowalski *et al.* [112]. The empty contours in the left and middle columns represent the constraints from the Gold sample [108, 109] and the ESSENCE data [110, 111], respectively. The right column shows the impact of the SCP nearby 1999 data. From Ref. [112].

After 1998 more SN Ia data have been collected by a number of high-redshift surveys – including SuperNova Legacy Survey [107] (SNLS), Hubble Space Telescope [108, 109] (HST), and “Equation of State: SupErNovae trace Cosmic Expansion” [110, 111] (ESSENCE) survey. The SNLS project, which is based on the Canada-France-Hawaii Telescope, consists of two components: (i) a large imaging survey to detect about 2000 supernovae and monitor their light curves, and (ii) a large spectroscopic survey to obtain supernovae identification and redshift. The HST survey is based on the image subtraction to search the SN Ia data in the high-redshift region $z > 1$ by including search depth, efficiency, timing, and false-positive discrimination. These data have been classified as the “Gold” data sets [109]. The ESSENCE project is a ground-based survey designed to detect about 200 SN Ia in the redshift range $z = 0.2\text{--}0.8$ to measure the equation of state of dark energy to better than 10%. In Fig. 5.4 the observational contours on $(\Omega_m^{(0)}, w_{\text{DE}})$ are plotted from the SN Ia data by Kowalski *et al.* [112]. Note that the equation of state of dark energy is assumed to be constant. While the SN Ia data alone are not yet sufficient to place tight bounds on w_{DE} , Fig. 5.4 clearly shows the presence of dark energy responsible for the late-time cosmic acceleration ($w_{\text{DE}} < -1/3$).

If the equation of state of dark energy varies in time, we need to parametrize w_{DE} as a function of the redshift z . If the SN Ia data are accurate enough to measure the luminosity distance $d_L(z)$ in terms of z , it is possible to determine the evolution of

w_{DE} by using Eqs. (2.86) and (2.88). However the error bars in the SN Ia data are still too large to determine $w_{\text{DE}}(z)$ without assuming some form of parametrization with respect to z . In particular, there is very little information on w_{DE} at $z > 0.5$ [112]. The parametrization of $w_{\text{DE}}(z)$ will be discussed in Section 7.5.

5.3 Cosmic Microwave Background

The observations of temperature anisotropies in the CMB provide another independent test for the existence of dark energy. The oldest sky we can see is the so-called last scattering surface at which electrons are trapped by hydrogen to form atoms (dubbed “decoupling” or “recombination”). The photons were tightly coupled to baryons and electrons before the decoupling epoch at $z \simeq 1090$, but they could freely move to us after that. In 1963 Penzias and Wilson [113] first detected the CMB photons thermalized to an almost uniform temperature across the sky. The temperature anisotropies of the CMB were first measured at large angular separations by the COBE satellite in 1992 [12]. The precise measurement of temperature anisotropies by high-precision experiments like BOOMERanG [114], MAXIMA [115], and especially WMAP [13] opened up a new opportunity to determine cosmological parameters to high precision.

All of the matter components in the Universe (dark matter, neutrinos, . . .) are coupled to gravity through the Einstein equations. The scalar part of the perturbations is the main source for the CMB temperature anisotropies. As we already mentioned the vector perturbations decay in the expanding Universe, whereas the tensor perturbations contribute to the CMB anisotropies as gravitational waves [90, 92]. However the amplitude of gravitational waves is suppressed relative to that of scalar perturbations if they originate from inflation (see Ref. [94]). Since the main feature of the CMB anisotropies is determined by scalar metric perturbations, we focus on scalar perturbations in the following discussion.

Numerical simulations are required to derive the complete spectra of temperature and polarization anisotropies because the Einstein and Boltzmann equations for photons as well as other matter components are coupled to each other. In spite of this complexity, Hu and Sugiyama [116, 117] obtained many fitting formulae which are very helpful to understand the CMB physics and also derived analytic solutions under some approximations. In the following we shall discuss how the presence of dark energy affects the CMB anisotropies. The first effect is the change of the position of acoustic peaks coming from the modification of the angular diameter distance. The second effect is the so-called integrated Sachs–Wolfe (ISW) effect caused by the variation of the gravitational potential [118] (see Section 4.11.1). Since the latter is limited to very large scales, the first effect is typically more important.

We have already derived in Section 4.9 the first-order equation for the direction-averaged (i.e. monopole) temperature anisotropy Θ_0 in the tight coupling regime ($\tau_{\text{op}} \gg 1$):

$$\Theta_0'' + \frac{R_s}{1 + R_s} \mathcal{H} \Theta_0' + k^2 c_s^2 \Theta_0 = -\frac{k^2}{3} \Psi - \frac{R_s}{1 + R_s} \mathcal{H} \Phi' - \Phi'', \quad (5.13)$$

where the sound speed squared is given by $c_s^2 = 1/3(1 + R_s)$ with the baryon-to-photon density ratio $R_s = 3\rho_b/4\rho_\gamma$. This ratio evolves as

$$R_s = \frac{3\omega_b}{4\omega_\gamma} \frac{1}{1 + z}, \quad (5.14)$$

where we have defined

$$\omega_b \equiv \Omega_b^{(0)} h^2, \quad \omega_\gamma \equiv \Omega_\gamma^{(0)} h^2. \quad (5.15)$$

If we take the value $\omega_b = 0.02267$ and $\omega_\gamma = 2.469 \times 10^{-5}$ we have $R_s = 2.7 \times 10^4/(1 + z)$, which is much smaller than unity for redshifts larger than 10^5 (corresponding to the regime of tight coupling between photons and baryons).

Taking the derivative of R_s with respect to η and using the definition of the redshift (2.28), we obtain the relation $R_s' = \mathcal{H} R_s$. Then Eq. (5.13) can be written as

$$\left[\frac{d^2}{d\eta^2} + \frac{R_s'}{1 + R_s} \frac{d}{d\eta} + k^2 c_s^2 \right] (\Theta_0 + \Phi) = \frac{k^2}{3} \left(\frac{1}{1 + R_s} \Phi - \Psi \right). \quad (5.16)$$

The second term on the l.h.s. of Eq. (5.16) is of the order of $(R_s/\eta^2)(\Theta_0 + \Phi) \approx R_s \mathcal{H}^2(\Theta_0 + \Phi)$, whereas the third term is of the order of $k^2 c_s^2(\Theta_0 + \Phi)$. Hence the second term can be neglected under the condition $R_s \ll c_s^2(k/\mathcal{H})^2$. In the tight-coupling regime ($R_s \ll 1$) this condition is well satisfied for the modes inside the Hubble radius ($k > \mathcal{H}$). Hence the homogeneous solution to Eq. (5.16) can be described by the sum of the solutions $\exp(\pm i \int k c_s d\eta)$, i.e.

$$(\Theta_0 + \Phi)^{(\text{hom})}(k, \eta) = c_1 f_1(k, \eta) + c_2 f_2(k, \eta), \quad (5.17)$$

where $f_1(\eta) = \sin[kr_s(\eta)]$, $f_2(\eta) = \cos[kr_s(\eta)]$, and r_s is the sound horizon defined by

$$r_s(\eta) \equiv \int_0^\eta d\tilde{\eta} c_s(\tilde{\eta}). \quad (5.18)$$

It is clear from Eq. (5.17) that the homogeneous solution leads to oscillations in CMB anisotropies with a time-dependent frequency $\omega_k(\eta) = kr_s(\eta)$.

The general solution to Eq. (5.16) can be obtained by the usual Green's function method. In the tight-coupling regime ($R_s \ll 1$) it is given by

$$(\Theta_0 + \Phi)(k, \eta) = c_1 f_1(\eta) + c_2 f_2(\eta) + \frac{k^2}{3} \int_0^\eta d\tilde{\eta} [\Phi(\tilde{\eta}) - \Psi(\tilde{\eta})] \frac{f_1(\tilde{\eta})f_2(\eta) - f_1(\eta)f_2(\tilde{\eta})}{f_1(\tilde{\eta})f_2'(\tilde{\eta}) - f_1'(\tilde{\eta})f_2(\tilde{\eta})}. \quad (5.19)$$

We take the initial conditions $\Theta'_0 = 0$ and $\Phi' = 0$ at $\eta = 0$, which determine the coefficients c_1 and c_2 . Then Eq. (5.19) reduces to

$$(\Theta_0 + \Phi)(k, \eta) = [\Theta_0(0) + \Phi(0)] \cos(kr_s) + \frac{k}{\sqrt{3}} \int_0^\eta d\tilde{\eta} [\Phi(\tilde{\eta}) - \Psi(\tilde{\eta})] \sin[k(r_s(\eta) - r_s(\tilde{\eta}))]. \quad (5.20)$$

Using Eqs. (4.176) and (5.20), the solution to the dipole moment is given by

$$\Theta_1(k, \eta) = \frac{1}{\sqrt{3}} [\Theta_0(0) + \Phi(0)] \sin(kr_s) - \frac{k}{3} \int_0^\eta d\tilde{\eta} [\Phi(\tilde{\eta}) - \Psi(\tilde{\eta})] \cos[k(r_s(\eta) - r_s(\tilde{\eta}))]. \quad (5.21)$$

The first term on the r.h.s. of Eq. (5.20) shows that there is a peak for the homogeneous solution of $\Theta_0 + \Phi$ at the position satisfying the relation

$$kr_s = n\pi, \quad (5.22)$$

where n are integers. The peak position is subject to change by the presence of the last term in Eq. (5.20). The dipole solution (5.21) also contributes to the CMB spectrum.

Recall that Eqs. (5.20) and (5.21) have been derived in the tight-coupling limit ($\tau_{\text{op}} \gg 1$). In order to obtain the CMB anisotropies observed today, we need to take into account the evolution of perturbations after the photons began to stream freely. (i.e. the epoch in which the tight coupling approximation is no longer valid). Moreover the contribution of higher-order moments (Θ_ℓ with $\ell \geq 2$) and the ISW effect modifies the shape of the power spectrum. Hu and Sugiyama [116] derived the following semi-analytic expression for the present temperature field (see also

[74] for the derivation):

$$\begin{aligned} \Theta_\ell(k, \eta_0) \simeq & [\Theta_0(k, \eta_*) + \Psi(k, \eta_*)] j_\ell[k(\eta_0 - \eta_*)] \\ & + 3\Theta_1(k, \eta_*) \left\{ j_{\ell-1}[k(\eta_0 - \eta_*)] - \frac{(\ell+1)j_\ell[k(\eta_0 - \eta_*)]}{k(\eta_0 - \eta_*)} \right\} \\ & + \int_0^{\eta_0} d\eta e^{-\tau_{\text{op}}} [\Psi'(k, \eta) - \Phi'(k, \eta)] j_\ell[k(\eta_0 - \eta)] , \end{aligned} \quad (5.23)$$

where $j_\ell(x)$ are spherical Bessel functions (see the Mathematical Appendix in Chapter 17) and η_* is the time at which the visibility function

$$g(\eta) \equiv -\tau'_{\text{op}} e^{-\tau_{\text{op}}} \quad (5.24)$$

takes a peak value. At the early stage much before the decoupling epoch (where the optical depth τ_{op} is much larger than 1), the visibility function $g(\eta)$ is nearly zero. It takes a peak value roughly around $z \sim 1000$. The scattering rate ($-\tau'_{\text{op}}$) decreases rapidly after the decoupling epoch and hence $g(\eta)$ approaches 0 again.

The analytic solution (5.23) reproduces numerical solutions within 10% accuracy. The first term on the r.h.s. of Eq. (5.23) comes from the integral $\int_0^{\eta_0} d\eta g(\eta) [\Theta_0(k, \eta) + \Psi(k, \eta)] j_\ell[k(\eta_0 - \eta)]$. This carries the information of the monopole solution (5.20) derived under the tight-coupling approximation. The effect of free streaming of photons appears in the visibility function $g(\eta)$ whose main contribution to the integral comes from the value at $\eta = \eta_*$ [note that $g(\eta)$ defined in Eq. (5.24) satisfies the normalization condition $\int_0^{\eta_0} d\eta g(\eta) = 1$]. The position of acoustic peaks estimated by Eq. (5.22) is shifted toward larger scales by the effect of free streaming of photons [74]. The second term in Eq. (5.23) coming from the dipole contribution Θ_1 also leads to the modification of the CMB power spectrum. The third term in Eq. (5.23) is responsible for the so-called ISW effect, which is induced by the variation of the gravitational potentials Ψ and Φ . While Ψ and Φ remain nearly constants during the matter-dominated epoch, the dominance of dark energy at late times gives rise to their variations. We have already discussed this issue in Section 4.11.1. In the limit $\ell \gg 1$ the spherical Bessel function $j_\ell(x)$ has a dependence $j_\ell(x) \simeq (1/\ell)(x/\ell)^{\ell-1/2}$, which is suppressed for large ℓ . Hence the dominant contribution to the ISW effect corresponds to the low ℓ modes ($\ell = \mathcal{O}(1)$).

When we confront the predicted temperature anisotropies with CMB observations, we expand the perturbation Θ in terms of spherical harmonics:

$$\Theta(\mathbf{x}, \eta) = \sum_{\ell=1}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m}(\mathbf{x}, \eta) Y_{\ell m}(\hat{n}) , \quad (5.25)$$

where the subscripts ℓ and m are conjugate to a real space unit vector $\hat{n}$ representing the direction of incoming photons. The spherical harmonics $Y_{\ell m}$ satisfies the

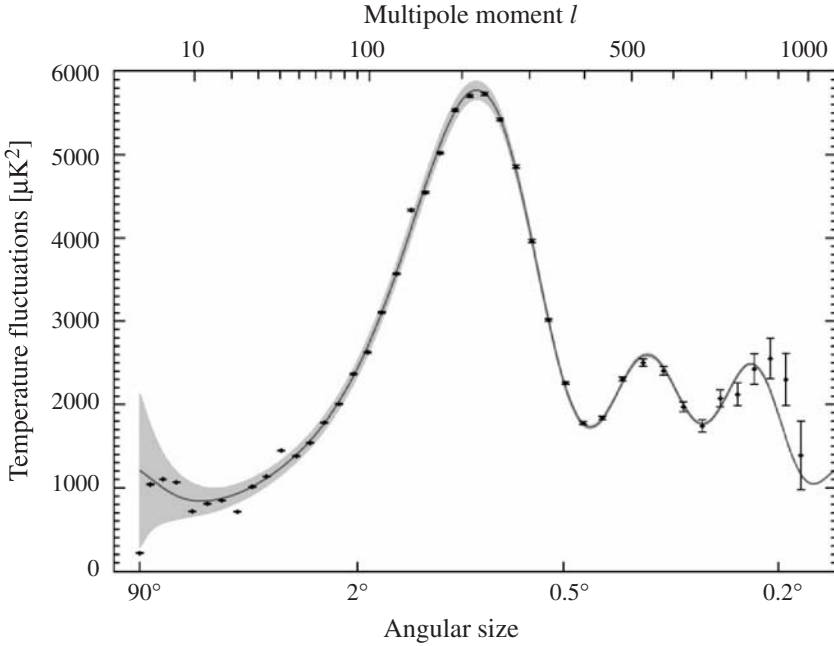


Figure 5.5 The CMB power spectrum $\ell(\ell + 1)C_\ell/2\pi$ versus the multiple moment ℓ and the angular size θ . The relation between ℓ and θ is $\theta = \pi/\ell$. The curve shows the theoretical prediction of the power spectrum, whereas the black points present the WMAP 5-year data. From the webpage of WMAP: <http://map.gsfc.nasa.gov/>

normalization condition

$$\int d\Omega Y_{\ell m}(\hat{n}) Y_{\ell' m'}^*(\hat{n}) = \delta_{\ell\ell'} \delta_{mm'}. \quad (5.26)$$

The coefficients $a_{\ell m}$ in Eq. (5.25) are assumed to be statistically independent. This means that the mean value of $a_{\ell m}$'s is zero ($\langle a_{\ell m} \rangle = 0$) with a non-zero variance defined by

$$C_\ell \equiv \langle |a_{\ell m}|^2 \rangle. \quad (5.27)$$

The variance C_ℓ can be expressed in terms of the temperature field $\Theta_\ell(k)$ in Fourier space (this equation will be derived in Section 14.1.1):

$$C_\ell = \frac{2}{\pi} \int_0^\infty dk k^2 |\Theta_\ell(k)|^2. \quad (5.28)$$

In Fig. 5.5 we show the predicted CMB temperature anisotropies $\ell(\ell + 1)C_\ell/2\pi$ versus the multipole moment ℓ together with the WMAP 5-year observational data [15]. The theoretical power spectrum agrees with the observational data with flying colors. The measured angle θ has a relation $\theta = \pi/\ell$ [rad] with the multipole ℓ .

Hence the larger scales correspond to lower values of ℓ . The large-scale power spectrum ($\ell \lesssim 10$) is dominated by the monopole mode Θ_0 , which inherits the information of nearly scale-invariant density perturbations generated during inflation. The reason why the quantity $\ell(\ell + 1)C_\ell/2\pi$ is plotted instead of C_ℓ itself is that the former is constant for scale-invariant perturbations on large scales, i.e. $\ell(\ell + 1)C_\ell = (\pi/2)\delta_H^2$, where δ_H^2 is the amplitude of curvature perturbations generated during inflation [74].

Equation (5.22) shows that the comoving wavelength corresponding to acoustic peaks can be approximately estimated as $\lambda_c = 2\pi/k = (2/n)r_s$. We then define the following characteristic angle for the location of peaks:

$$\theta_A \equiv \frac{r_s(z_{\text{dec}})}{d_A^{(c)}(z_{\text{dec}})}, \quad (5.29)$$

where z_{dec} is the redshift at the decoupling epoch and $d_A^{(c)}$ is the comoving angular diameter distance defined by

$$d_A^{(c)}(z) \equiv \frac{d_A(z)}{a} = (1+z)d_A(z). \quad (5.30)$$

The physical (proper) diameter distance $d_A(z)$ is given by Eq. (2.73). The multipole ℓ corresponding to the angle (5.29) is

$$\ell_A = \frac{\pi}{\theta_A} = \pi \frac{d_A^{(c)}(z_{\text{dec}})}{r_s(z_{\text{dec}})}. \quad (5.31)$$

From Eqs. (2.73) and (5.30) the comoving angular diameter distance $d_A^{(c)}(z_{\text{dec}})$ is expressed as

$$d_A^{(c)}(z_{\text{dec}}) = \frac{c}{H_0} \frac{1}{\sqrt{\Omega_m^{(0)}}} \mathcal{R}, \quad (5.32)$$

where $\mathcal{R}$ is the so-called CMB shift parameter defined by

$$\mathcal{R} = \sqrt{\frac{\Omega_m^{(0)}}{\Omega_K^{(0)}}} \sinh \left(\sqrt{\Omega_K^{(0)}} \int_0^{z_{\text{dec}}} \frac{dz}{E(z)} \right). \quad (5.33)$$

From Eqs. (4.180) and (5.18), the sound horizon $r_s(z_{\text{dec}})$ is

$$r_s(z_{\text{dec}}) = \frac{c}{\sqrt{3}a_0H_0} \int_{z_{\text{dec}}}^{\infty} \frac{dz}{\sqrt{1+R_s} E(z)}. \quad (5.34)$$

For the redshift z_{dec} there is a fitting formula by Hu and Sugiyama [117]

$$z_{\text{dec}} = 1048 \left(1 + 0.00124\omega_b^{-0.738} \right) \left(1 + g_1\omega_m^{g_2} \right), \quad (5.35)$$

where $\omega_m \equiv \Omega_m^{(0)} h^2$ and

$$g_1 = 0.0783 \omega_b^{-0.238} / (1 + 39.5 \omega_b^{0.763}), \quad g_2 = 0.560 / (1 + 21.1 \omega_b^{1.81}). \quad (5.36)$$

The bounds coming from the WMAP 5-year data correspond to $\omega_b = 0.02265 \pm 0.00059$ and $\omega_m = 0.1369 \pm 0.0037$. Taking the values $\omega_b = 0.02265$ and $\omega_m = 0.1369$, we obtain $z_{\text{dec}} = 1090.98$ from Eq. (5.35).

Since the contribution of dark energy to $E(z)$ in Eq. (5.34) is negligible for $z > z_{\text{dec}}$, one can estimate this quantity to be $E = (\sqrt{a + a_{\text{eq}}}/a^2) \sqrt{\Omega_m^{(0)}}$, where $a_{\text{eq}} = (1 + z_{\text{eq}})^{-1}$ is the scale factor at the radiation–matter equality [see Eq. (2.78)]. Then the integral (5.34) can be written as

$$r_s(z_{\text{dec}}) = \frac{c}{\sqrt{3} H_0} \frac{1}{\sqrt{\Omega_m^{(0)}}} \int_0^{a_{\text{dec}}} \frac{1}{\sqrt{1 + R_s(a)}} \frac{1}{\sqrt{a + a_{\text{eq}}}} da, \quad (5.37)$$

where

$$R_s(a) = (3\omega_b/4\omega_\gamma) a. \quad (5.38)$$

Equation (5.37) is integrated to give

$$r_s(z_{\text{dec}}) = \frac{4}{3} \frac{ch}{H_0} \sqrt{\frac{\omega_\gamma}{\omega_m \omega_b}} \ln \left(\frac{\sqrt{R_s^{(\text{dec})} + R_s^{(\text{eq})}} + \sqrt{1 + R_s^{(\text{dec})}}}{1 + \sqrt{R_s^{(\text{eq})}}} \right), \quad (5.39)$$

where $R_s^{(\text{dec})} \equiv R_s(a_{\text{dec}})$ and $R_s^{(\text{eq})} \equiv R(a_{\text{eq}})$.

From Eqs. (5.32) and (5.39) the multipole ℓ_A in Eq. (5.31) is

$$\ell_A = \frac{3\pi}{4} \sqrt{\frac{\omega_b}{\omega_\gamma}} \mathcal{R} \left[\ln \left(\frac{\sqrt{R_s^{(\text{dec})} + R_s^{(\text{eq})}} + \sqrt{1 + R_s^{(\text{dec})}}}{1 + \sqrt{R_s^{(\text{eq})}}} \right) \right]^{-1}, \quad (5.40)$$

which shows that ℓ_A is proportional to $\mathcal{R}$. The CMB shift parameter defined in Eq. (5.33) is affected by the cosmic expansion history from the decoupling to the present. The presence of dark energy leads to a shift of $\mathcal{R}$ compared to the CDM model, thereby changing the value of ℓ_A . Hence the CMB shift parameter can be used to place constraints on dark energy. The bound on $\mathcal{R}$ according to the WMAP 5-year data is given by

$$\mathcal{R} = 1.710 \pm 0.019 \quad (68\% \text{ confidence level}). \quad (5.41)$$

For example, let us compute the multipole ℓ_A for $\mathcal{R} = 1.710$, $\omega_b = 0.02265$, $\omega_m = 0.1369$, and $\omega_\gamma = 2.469 \times 10^{-5}$. Using Eqs. (2.77), (5.35), and (5.38) we obtain the value $\ell_A = 299$. This is different from the location of the first acoustic peak

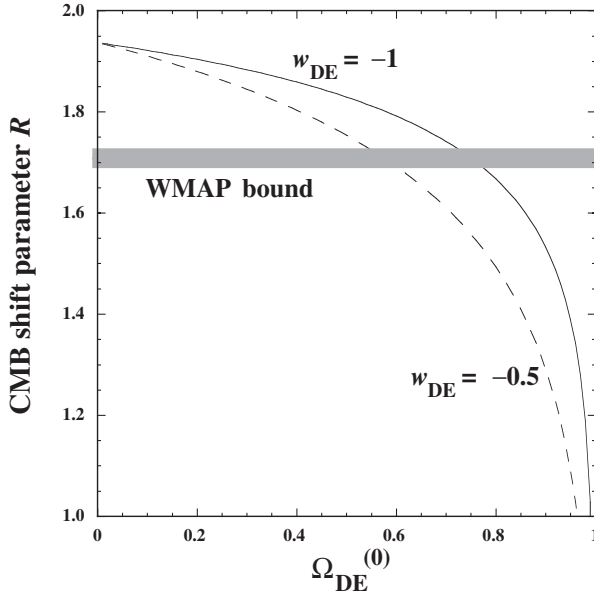


Figure 5.6 The CMB shift parameter $\mathcal{R}$ versus $\Omega_{\text{DE}}^{(0)}$ for $w_{\text{DE}} = -1$ and $w_{\text{DE}} = -0.5$ in the flat Universe. We also show the bound coming from the WMAP 5-year observational data. For the cosmological constant ($w_{\text{DE}} = -1$) the observational constraint on $\Omega_{\text{DE}}^{(0)}$ coming from the CMB shift parameter is $0.72 < \Omega_{\text{DE}}^{(0)} < 0.77$.

which is located around $\ell_1 = 220$ (see Fig. 5.5). As we already explained, this shift comes from several effects such as the free streaming of photons and the dipole contribution. We write the general relation for all peaks and troughs of observed CMB anisotropies as

$$\ell_m = \ell_A(m - \phi_m), \quad (5.42)$$

where m represents peak numbers ($m = 1$ for the first peak, $m = 1.5$ for the first trough, ...) and ϕ_m is the shift of multipoles. It is known that ϕ_m depends weakly on ω_b and ω_m for a given cosmic curvature $\Omega_K^{(0)}$. According to the fits by Doran and Lilley [119] the shift of the first peak is about $\phi_1 = 0.265$, which gives $\ell_1 = 220$ for $\ell_A = 299$.

In the flat Universe the CMB shift parameter (5.33) reduces to

$$\mathcal{R} = \sqrt{\Omega_m^{(0)}} \int_0^{z_{\text{dec}}} \frac{dz}{E(z)}, \quad (5.43)$$

where $E(z) = [\Omega_m^{(0)}(1+z)^3 + \Omega_r^{(0)}(1+z)^4 + \Omega_{\text{DE}}^{(0)}(1+z)^{3(1+w_{\text{DE}})}]^{1/2}$ for the constant dark energy equation of state. In Fig. 5.6 we plot $\mathcal{R}$ versus $\Omega_{\text{DE}}^{(0)}$ for the cases $w_{\text{DE}} = -1$ and $w_{\text{DE}} = -0.5$ together with the WMAP bound (5.41). For larger

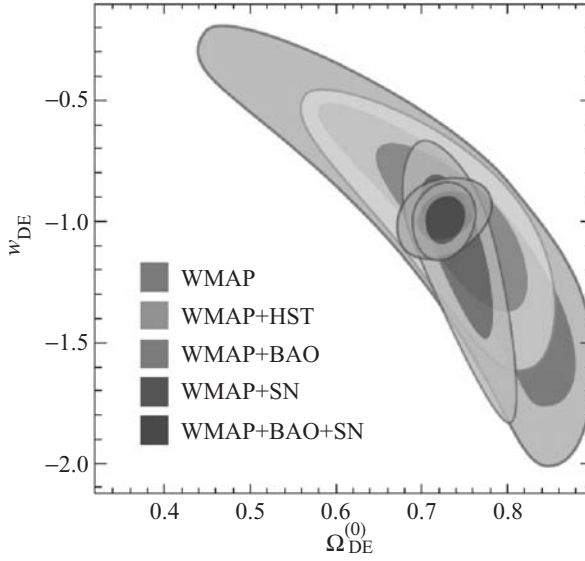


Figure 5.7 Observational contours (68% and 95% confidence level) on the constant dark energy equation of state w_{DE} and the present dark energy density $\Omega_{\text{DE}}^{(0)}$ for the flat Universe. The WMAP 5-year data alone do not provide strong constraints on w_{DE} and $\Omega_{\text{DE}}^{(0)}$. This is slightly improved by adding the measurement of H_0 from the Hubble key result. The observational contours coming from WMAP+BAO, WMAP+SN Ia, and WMAP+BAO+SN Ia are also shown. The combined analysis of WMAP+SN Ia data provide a tight constraint $-1.098 < w_{\text{DE}} < -0.841$ at the 95% confidence level. From Ref. [15].

$\Omega_{\text{DE}}^{(0)}$ the CMB shift parameter (5.43) gets smaller. When $w_{\text{DE}} = -1$ the density parameter is constrained to be $0.72 < \Omega_{\text{DE}}^{(0)} < 0.77$ from the bound (5.41). From Fig. 5.6 we find that the observationally allowed values of $\Omega_{\text{DE}}^{(0)}$ become smaller for increasing w_{DE} . Since the CMB shift parameter depends only weakly on w_{DE} , the equation of state w_{DE} is not strongly constrained by the CMB data alone.

In Fig. 5.7 the combined observational bounds on $\Omega_{\text{DE}}^{(0)}$ and (constant) w_{DE} are plotted in the flat Universe. This is the joint analysis derived by using the observational data of WMAP 5-year, SN Ia, and BAO. As expected, the WMAP 5-year data alone do not provide strong constraints on w_{DE} , although the evidence for the accelerated expansion ($w_{\text{DE}} < -1/3$) can be seen in Fig. 5.7. Notice that the WMAP bound in Fig. 5.7 is consistent with the bound coming from the CMB shift parameter discussed above. For decreasing w_{DE} the observationally allowed range of $\Omega_{\text{DE}}^{(0)}$ shifts to larger values. The prior on the Hubble constant $H_0 = 72 \pm 8 \text{ km sec}^{-1} \text{ Mpc}^{-1}$ from the Hubble Key Project [72] improves the constraint a bit: $-1.47 < w_{\text{DE}} < -0.58$ at the 95% confidence level. Adding the

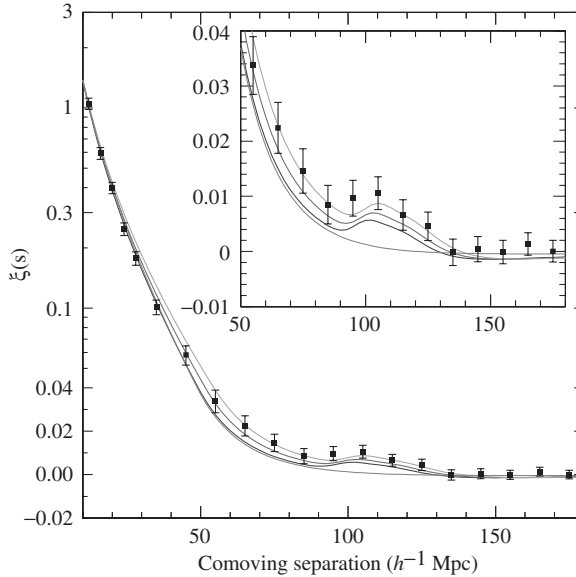


Figure 5.8 The large-scale redshift-space correlation function of the SDSS sample. The inset shows an expanded view with a linear vertical axis. From top to bottom each curve corresponds to $\Omega_m^{(0)}h^2 = 0.12, 0.13, 0.14$ with $\Omega_b^{(0)}h^2 = 0.024$ and a pure CDM (no baryons) model with $\Omega_m^{(0)}h^2 = 0.15$. The observational data clearly show the existence of an acoustic peak around the comoving separation scale $100 h^{-1}$ Mpc, in agreement with the predictions except for the pure CDM model. From Ref. [68].

“Union” SN Ia data by Kowalski *et al.* [112] significantly improves the constraint: $-1.098 < w_{\text{DE}} < -0.841$ at the 95% confidence level. This shows the importance of combined analysis using independent observational data. In the next subsection we shall proceed to the constraint coming from baryon acoustic oscillations.

5.4 Baryon acoustic oscillations

Since baryons are tightly coupled to photons before the recombination epoch, the oscillations of sound waves should be imprinted in the baryon perturbations as well as the CMB temperature anisotropies. Eisenstein *et al.* [68] found a peak of baryon acoustic oscillations in the large-scale correlation function at $100 h^{-1}$ Mpc separation measured from a spectroscopic sample of 46,748 luminous red galaxies from the Sloan Digital Sky Survey (SDSS), see Fig. 5.8. This detection of baryon oscillations provided another independent test for constraining the property of dark energy, which contributes to breaking residual degeneracies in the CMB data [121].

At the tight coupling epoch in which baryons are strongly coupled to photons the perturbations in baryons, $\delta_b \equiv \delta\rho_b/\rho_b$, satisfy the following equation of motion

$$\delta_b'' + \frac{R_s}{1 + R_s} \mathcal{H} \delta_b' + k^2 c_s^2 \delta_b = -k^2 \Psi - \frac{3R_s}{1 + R_s} \mathcal{H} \Phi' - 3\Phi'' . \quad (5.44)$$

If we substitute the relation $\delta_b = 3\Theta_0$ into Eq. (5.44), we obtain the same equation as Eq. (5.13) for the temperature field Θ_0 . Since $\Theta_0 = (1/4)\delta_\gamma$, this corresponds to

$$\delta_b = \frac{3}{4} \delta_\gamma , \quad (5.45)$$

which is the adiabatic condition we have already mentioned after Eq. (4.198). This is also associated with the fact that the ratio n_b/s of the number density of baryons (n_b) to the entropy density ($s = (\rho + P)/T$) does not vary in time. During the radiation era the entropy density is dominated by the contribution of relativistic particles, which gives $s \propto T^3 \propto \rho_\gamma^{3/4}$. Using the relation $\rho_b \propto n_b$, we find

$$\frac{\delta(n_b/s)}{n_b/s} = \frac{\delta n_b}{n_b} - \frac{\delta s}{s} = \delta_b - \frac{3}{4} \delta_\gamma , \quad (5.46)$$

which vanishes under the condition (5.45). Since the baryon–entropy ratio is independent of the temperature T , it takes a constant value at different spatial positions when the perturbations are generated. Hence the condition $\delta_b = (3/4)\delta_\gamma$ sets the (adiabatic) initial condition for baryon perturbations. This relation is preserved in the tight-coupling era so that baryons and photons evolve as a single fluid.

The sound horizon at which baryons were released from the Compton drag of photons plays a crucial role to determine the location of baryon acoustic oscillations. This epoch, called the *drag epoch*, occurs at the redshift z_{drag} . The sound horizon at $z = z_{\text{drag}}$ is

$$r_s(z_{\text{drag}}) = \int_0^{\eta_{\text{drag}}} d\eta c_s(\eta) , \quad (5.47)$$

where c_s is given in Eq. (4.180). We caution that the drag epoch does not coincide with the recombination epoch at which photons were released from electrons. For the redshift z_{drag} there is a fitting formula by Eisenstein and Hu [120]:

$$z_{\text{drag}} = \frac{1291 \omega_m^{0.251}}{1 + 0.659 \omega_m^{0.828}} (1 + b_1 \omega_b^{b_2}) , \quad (5.48)$$

where

$$b_1 = 0.313 \omega_m^{-0.419} (1 + 0.607 \omega_m^{0.674}) , \quad b_2 = 0.238 \omega_m^{0.223} . \quad (5.49)$$

The WMAP 5-year data constrain the values of z_{drag} and $r_s(z_{\text{drag}})$ to be $z_{\text{drag}} = 1020.5 \pm 1.6$ and $r_s(z_{\text{drag}}) = 153.3 \pm 2.0$ Mpc [15].

We observe the angular and redshift distributions of galaxies as a power spectrum $P(k_{\perp}, k_{\parallel})$ in the redshift space, where $k_{\perp}$ and $k_{\parallel}$ are the wavenumbers perpendicular and parallel to the direction of light respectively. As we will show in more detail in Section 14.2, it is possible to measure the following two ratios [122]

$$\theta_s(z) = \frac{r_s(z_{\text{drag}})}{(1+z)d_A(z)}, \quad (5.50)$$

$$\delta z_s(z) = \frac{r_s(z_{\text{drag}})H(z)}{c}, \quad (5.51)$$

where the denominator in Eq. (5.50) corresponds to the comoving diameter distance $d_A^{(c)}(z) = (1+z)d_A(z)$ defined in Eq. (5.30). The angle $\theta_s(z)$ is completely analogous to the CMB acoustic peak angle introduced in Eq. (5.29) and corresponds to observations orthogonal to the line of sight. The quantity δz_s is instead measured by identifying in the fluctuation spectrum the oscillations along the line of sight.

So far the BAO data have not been accumulated sufficiently to measure the two distances $\theta_s(z)$ and $\delta z_s(z)$ separately [123]. However it is possible to obtain a combined distance scale ratio from the spherically averaged spectrum:

$$[\theta_s(z)^2 \delta z_s(z)]^{1/3} \equiv \frac{r_s(z_{\text{drag}})}{[(1+z)^2 d_A^2(z) c / H(z)]^{1/3}}, \quad (5.52)$$

or the related effective distance [68]

$$D_V(z) \equiv \left[(1+z)^2 d_A^2(z) \frac{cz}{H(z)} \right]^{1/3}, \quad (5.53)$$

obtained from the combination of two spatial dimensions orthogonal to the direction of sight and one dimension along the direction of sight. The current constraint from SDSS data is $D_V(z = 0.35) = 1370 \pm 64$ Mpc at a typical redshift $z = 0.35$ [68].

Another observational constraint comes from the 2-degree Field (2dF) Galaxy Redshift Survey. Defining the relative BAO distance

$$r_{\text{BAO}}(z) \equiv r_s(z_{\text{drag}}) / D_V(z), \quad (5.54)$$

we have at the two redshifts ($z = 0.2$ and $z = 0.35$) [124]

$$r_{\text{BAO}}(z = 0.2) = 0.1980 \pm 0.0058, \quad r_{\text{BAO}}(z = 0.35) = 0.1094 \pm 0.0033. \quad (5.55)$$

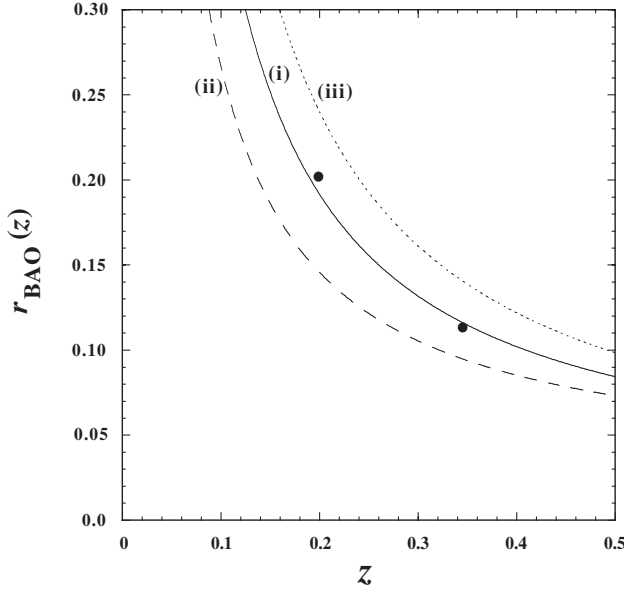


Figure 5.9 The BAO distance ratio (5.54) versus the redshift z in the flat Universe ($K = 0$) with the cosmological constant ($w_{\text{DE}} = -1$) for three different cases: (i) $\Omega_{\text{DE}}^{(0)} = 0.75$, (ii) $\Omega_{\text{DE}}^{(0)} = 0$, and (iii) $\Omega_{\text{DE}}^{(0)} = 0.95$. We choose the values $\omega_\gamma = 2.469 \times 10^{-5}$ and $\omega_b = 0.02265$. We also show the most likely observational values of $r_{\text{BAO}}(z)$ at $z = 0.2$ and $z = 0.35$ as black points. The observational data favor the dark energy density around $\Omega_{\text{DE}}^{(0)} = 0.75$.

The sound horizon $r_s(z_{\text{drag}})$ can be obtained by replacing $R_s^{(\text{dec})}$ in Eq. (5.39) for $R_s^{(\text{drag})} \equiv R_s(z_{\text{drag}})$, whereas the angular diameter distance d_A is given in Eq. (2.73). This gives the explicit form for $r_{\text{BAO}}(z)$:

$$r_{\text{BAO}}(z) = \frac{4}{3} \sqrt{\frac{\omega_\gamma}{\Omega_m^{(0)} \omega_b}} \left[\frac{z}{E(z)} \right]^{-1/3} \left[\frac{1}{\sqrt{\Omega_K^{(0)}}} \sinh \left(\sqrt{\Omega_K^{(0)}} \int_0^z \frac{d\tilde{z}}{E(\tilde{z})} \right) \right]^{-2/3} \times \ln \left(\frac{\sqrt{R_s^{(\text{drag})} + R_s^{(\text{eq})}} + \sqrt{1 + R_s^{(\text{drag})}}}{1 + \sqrt{R_s^{(\text{eq})}}} \right). \quad (5.56)$$

In Fig. 5.9 we plot $r_{\text{BAO}}(z)$ versus z for the Λ CDM model in the flat Universe ($w_{\text{DE}} = -1$ and $K = 0$). This shows that the Λ CDM model with $\Omega_{\text{DE}}^{(0)} = 0.75$ is favored over the CDM model from the BAO data at $z = 0.2$ and $z = 0.35$.

If the equation of state of dark energy w_{DE} is not -1 but a constant, the models with $w_{\text{DE}} < -1$, dubbed “phantoms,” are favored by the present BAO data. This

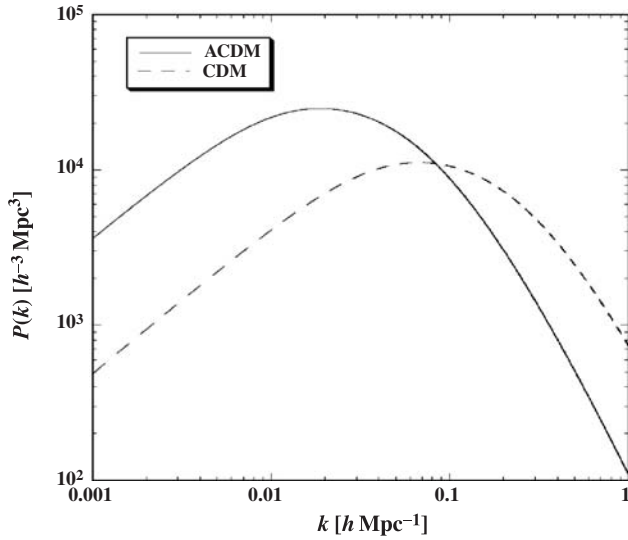


Figure 5.10 The predicted matter power spectra P_{δ_m} in the flat Universe for the Λ CDM model with $\Omega_m^{(0)} = 0.28$ and the CDM model with $\Omega_m^{(0)} = 1.0$. We use the expression (4.212) of the power spectrum with model parameters $\delta_H^2 = 3.2 \times 10^{-10}$, $n_s = 0.96$, and $h = 0.7$. For smaller $\Omega_m^{(0)}$ the position of the peak is shifted toward smaller k .

can be seen in the combined analysis of WMAP 5-year and BAO data in Fig. 5.7, which shows that even the equation of state such as $w_{\text{DE}} = -1.5$ can be allowed. However the addition of the SN Ia data provides a tight constraint: $-1.097 < w_{\text{DE}} < -0.858$ at the 95% confidence level. This constraint is not much different from the joint analysis of the WMAP+SN Ia data without the BAO data ($-1.098 < w_{\text{DE}} < -0.841$ at the 95% confidence level).

The non-linear evolution of baryon acoustic oscillations was extensively studied in Refs. [125, 126] by comparing N -body simulations with analytically estimated spectra for the Λ CDM model.

5.5 Large-scale structure

The observations of large-scale structure such as the galaxy clustering provides another test for the existence of dark energy. The wavenumber at the peak position of P_{δ_m} corresponds to k_{eq} given in Eq. (4.195). This shows that k_{eq} decreases for smaller values of $\Omega_m^{(0)}$. In Fig. 5.10 we plot the predicted matter power spectrum for the flat Λ CDM model ($\Omega_m^{(0)} = 0.28$) and for the CDM model ($\Omega_m^{(0)} = 1.0$). In the presence of dark energy the peak position shifts toward larger scales (i.e. for smaller

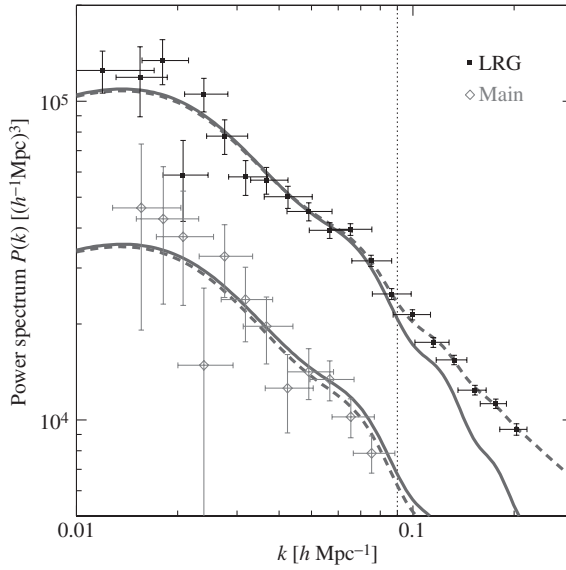


Figure 5.11 Measured power spectra with error bars for the full luminous red galaxies (LRG) and main galaxy samples of the 2dF survey. The solid curves show the theoretically predicted spectra for the Λ CDM model obtained in the linear perturbation theory with galaxy bias $b = 1.9$ (top) and $b = 1.1$ (bottom) relative to the $z = 0$ matter power spectrum. The dashed curves include the non-linear correction to the matter spectrum by Cole *et al.* [88]. The non-linear effect becomes important for the scales $k \gtrsim 0.09 h \text{ Mpc}^{-1}$. From Ref. [127].

k). For $\Omega_m^{(0)} = 1$ and $\Omega_m^{(0)} = 0.28$ with $h = 0.7$ we have $k_{\text{eq}} = 0.051 h \text{ Mpc}^{-1}$ and $k_{\text{eq}} = 0.014 h \text{ Mpc}^{-1}$, respectively. Hence the scale of the peak position can be used as a probe of dark energy.

As we have seen in Section 4.8 the matter spectrum P_{δ_m} is related to the observed galaxy power spectrum P_{δ_g} via the relation

$$P_{\delta_g} = b^2 P_{\delta_m}, \quad (5.57)$$

where b is the bias parameter. In Fig. 5.11 the galaxy power spectra of luminous red galaxies (LRG) and main galaxy samples of the SDSS are plotted [127]. The position of the peak, around the scale $0.01 h \text{ Mpc}^{-1} < k < 0.02 h \text{ Mpc}^{-1}$, shows that the Λ CDM model is favored over the CDM model. Although the galaxy power spectra alone do not provide tight bounds on the density parameter $\Omega_{\text{DE}}^{(0)}$ as well as w_{DE} , the important point is that the observations of LSS are consistent with the existence of dark energy. Notice that the linear spectrum plotted in Fig. 5.10

is modified for the scales $k \gtrsim 0.09 h \text{ Mpc}^{-1}$ because non-linear effects become important on smaller scales [88].

5.6 Problems

- 5.1** Integrate Eq. (5.12) numerically for four different values $\Omega_{\text{DE}}^{(0)} = 0, 0.3, 0.7, 1$ and plot the luminosity distance d_L (in units of cH_0^{-1}) in the flat Universe as a function of z .
- 5.2** Plot the BAO distance ratio (5.56) versus the redshift z for $\Omega_{\text{DE}}^{(0)} = 0$, $\Omega_{\text{DE}}^{(0)} = 0.75$, and $\Omega_{\text{DE}}^{(0)} = 0.95$.

6

Cosmological constant

The simplest candidate for dark energy is the cosmological constant Λ , which is so called because its energy density is constant in time and space. In fact the Λ CDM model has been systematically proved consistent with a large number of observations. The Lagrangian density for the Λ CDM model is simply given by the linear term in R plus Λ , see Eq. (6.2). Despite its simplicity it is generally difficult to explain why the energy scale of the cosmological constant required for the cosmic acceleration today is very small relative to that predicted by particle physics. As we already mentioned, the vacuum energy density evaluated by summing the zero-point energy of a scalar field is about 10^{121} times larger than the observed dark energy density (for a momentum cut-off around the Planck scale).

The problem of a large value of Λ was present long before the observational discovery of the late-time cosmic acceleration. In fact, even if we had no observational evidence of dark energy we would still need to understand why the cosmological constant vanishes. Models of dark energy alternative to Λ CDM are based on the assumption that Λ is zero or negligible. So the problem of the cosmological constant is to find some mechanism that either makes it vanish or renders it a very small value compatible with the present cosmological density. In the former case the origin of dark energy needs to be explored further, but in the latter case the problems of the cosmological constant and dark energy are solved at the same time.

In this chapter we first present the action for the Λ CDM model and show that the Einstein equations follow from this by the action principle. We then proceed to the history of the cosmological constant and its fine-tuning problem. We review a number of attempts to solve the cosmological constant problem in the framework of supergravity and superstring theories. We also discuss several topics related to the cosmological constant – such as the anthropic selection, and the decoupling of Λ from gravity.

6.1 Einstein equations with the cosmological constant

The energy-momentum tensor $T_{\mu\nu}$ on the r.h.s. of the Einstein equations (2.8) obeys the conservation law $T_{\mu\nu}{}^{;\nu} = 0$. Since the metric $g_{\mu\nu}$ satisfies the relation $g_{\mu\nu}{}^{;\nu} = 0$, it is possible to add the term $\Lambda g_{\mu\nu}$ to the Einstein equations:

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}, \quad (6.1)$$

where Λ is the cosmological constant. It is interesting to note that these are the most general equations of second-order in the metric in four dimensions. In scalar-tensor metric theories an additional term coupled to a Gauss–Bonnet term is also allowed.

The Einstein equations (6.1) can be derived by the action principle. It is based on the linear action in terms of the Ricci scalar $R = g^{\mu\nu} R_{\mu\nu}$ and the matter action S_m :

$$S = \frac{1}{16\pi G} \int d^4x \sqrt{-g} (R - 2\Lambda) + S_m. \quad (6.2)$$

Let us derive the Einstein equations (6.1) from the action (6.2). It is important to understand this procedure because similar steps are taken when we consider modified gravity models in Chapter 9.

The variation of the action (6.2) with respect to $g^{\mu\nu}$ gives

$$\begin{aligned} \delta S = \frac{1}{16\pi G} \int d^4x \left[\delta(\sqrt{-g})(g^{\mu\nu} R_{\mu\nu} - 2\Lambda) \right. \\ \left. + \sqrt{-g} \delta g^{\mu\nu} R_{\mu\nu} + \sqrt{-g} g^{\mu\nu} \delta R_{\mu\nu} \right] + \delta S_m. \end{aligned} \quad (6.3)$$

Since $\delta R_{\mu\nu} = (\delta \Gamma_{\mu\nu}^\alpha)_{;\alpha} - (\delta \Gamma_{\mu\alpha}^\alpha)_{;\nu}$ we have $g^{\mu\nu} \delta R_{\mu\nu} = (g^{\mu\nu} \delta \Gamma_{\mu\nu}^\alpha - g^{\mu\alpha} \delta \Gamma_{\mu\nu}^\nu)_{;\alpha}$ and hence

$$\int d^4x \sqrt{-g} g^{\mu\nu} \delta R_{\mu\nu} = \int d^4x \sqrt{-g} (g^{\mu\nu} \delta \Gamma_{\mu\nu}^\alpha - g^{\mu\alpha} \delta \Gamma_{\mu\nu}^\nu)_{;\alpha} = 0, \quad (6.4)$$

where we have employed Gauss's theorem in the last equality. This shows that the last term in the square bracket of Eq. (6.3) vanishes. Now we also use the relation $\delta(\sqrt{-g}) = -(1/2)\sqrt{-g} g_{\mu\nu} \delta g^{\mu\nu}$. This can easily be derived by differentiating with respect to $g^{\mu\nu}$ the determinant g written as $g_{\mu\nu} \mathcal{M}^{(\mu\nu)}$ where $\mathcal{M}^{(\mu\nu)}$ is the determinant of the cofactor matrix, which does not depend on the element $g_{\mu\nu}$ itself, and then replacing $\mathcal{M}^{(\mu\nu)} = g g^{\mu\nu}$. Then Eq. (6.3) reads

$$\delta S = \frac{1}{16\pi G} \int d^4x \sqrt{-g} \left(R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu} \right) \delta g^{\mu\nu} + \delta S_m. \quad (6.5)$$

The energy-momentum tensor $T_{\mu\nu}$ is defined from the variation of δS_m in terms of $g^{\mu\nu}$:

$$\delta S_m = -\frac{1}{2} \int d^4x \sqrt{-g} T_{\mu\nu} \delta g^{\mu\nu}. \quad (6.6)$$

Then Eq. (6.5) reduces to

$$\delta S = \frac{1}{16\pi G} \int d^4x \sqrt{-g} \left(R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu} - 8\pi G T_{\mu\nu} \right) \delta g^{\mu\nu}. \quad (6.7)$$

The Einstein equations (6.1) follow from the action principle, $\delta S = 0$.

6.2 History of the cosmological constant

After Einstein constructed General Relativity in 1915–1916 [128], he tried to apply his theory to the Universe in 1917 [23]. In the absence of the cosmological constant it is obvious from Eq. (2.19) that the scale factor a can dynamically change in time (except in the case of a fluid at rest with a specific equation of state: $w = P/\rho = -1/3$). In the 1910s, however, Einstein believed that the Universe was static and introduced the cosmological constant to realize such a Universe.

For the FLRW metric (2.1) the Einstein equations (6.1) read

$$H^2 = \frac{8\pi G}{3} \rho - \frac{K}{a^2} + \frac{\Lambda}{3}, \quad (6.8)$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3}. \quad (6.9)$$

From Eq. (6.9) it is clear that Λ works as a repulsive force against gravity at the background level. In the Universe dominated by a pressureless matter ($P = 0$), we find that the static Universe ($\dot{a} = \ddot{a} = 0$) corresponds to

$$\rho = \frac{\Lambda}{4\pi G}, \quad \frac{K}{a^2} = \Lambda. \quad (6.10)$$

This equation, the first relativistic cosmology ever, shows that the density ρ in the Universe is determined by Λ . Einstein believed that this solution (a “crazy idea” according to his own words in a letter to de Sitter) was a way to embody Mach’s idea of linking mass (ρ) to inertia, here represented by space-time geometry $g_{\mu\nu}$. He thought that he could eventually show that matter was necessary to define a non-Minkowskian metric. However, the above static solution is unstable against perturbations of the density ρ as was later demonstrated by Lemaître. In fact, if $\Lambda/3 > (4\pi G\rho)/3$, Eq. (6.9) shows that the Universe departs from the static point given in Eq. (6.10) with the growth of a . If $\Lambda/3 < (4\pi G\rho)/3$ the Universe is also away from the static point with the decrease of a . Einstein did not realize

this instability since he did not write down the differential equation for $a(t)$. Shortly after, in the same year 1917, de Sitter [129] found his accelerated solution $H = \sqrt{\Lambda/3}$ in empty space, paving the way to the dismissal of Mach's principle in cosmology.

At the same time, from 1910 to the mid 1920s, Slipher was observing the spectra of galaxies (spiral nebulae) and found most of them to be red-shifted. In 1922 Friedmann found the evolving solution that represents the expanding Universe [130]. In 1927 Lemaître [131] studied the relation between the observed results of the redshift and the homogeneous Universe dominated by a pressureless dust. In Lemaître's model there are three distinct periods for the evolution of the Universe:

- (i) A period of cosmic expansion ($a \propto t^n$ with $0 < n < 1$) from a point source during which the basic elements were formed. This corresponds to the expanding Universe dominated by matter (either the pressureless matter or the radiation).
- (ii) A period of a very slow expansion ($a \propto \text{constant}$) during which nebulae were formed. This resembles the static Universe proposed by Einstein.
- (iii) A period of a fast expansion ($a \propto t^n$ with $n > 1$) during which the recession of the nebulae is accelerating. This period can be realized by the de Sitter solution ($H = \sqrt{\Lambda/3}$) in the presence of the cosmological constant.

Lemaître's model is the first "hot Big Bang" model, in which the matter density ρ goes to infinity as $a \rightarrow 0$. Apart from the existence of the period (ii) Lemaître's model describes well the evolution of the Universe even in the modern context. The loitering period (ii) should be replaced by a short transient period from the matter era to the accelerated epoch during which the system crosses the point $\ddot{a} = 0$, while the nebulae formed during the matter-dominated epoch [the period (i)]. The period (iii) is exactly the phase of the late-time cosmic acceleration realized by the presence of Λ . We can say that Lemaître, influenced by de Sitter's accelerated solution, produced the first consistent dark energy model.

In 1929 Hubble formulated Hubble's law (2.32) by combining his measurements of galaxy distances with Slipher's measurements of the redshifts associated with the galaxies [71]. This was the first direct quantitative evidence for the expansion of the Universe. The existence of the cosmological constant was clearly not required to give rise to a (decelerated) cosmic expansion. In the book "The Meaning of Relativity" written by Einstein in 1945 [132], he stated that "if Hubble's expansion had been discovered at the time of the creation of the general theory of relativity, the cosmological member (the cosmological constant) would never have been introduced." In 1970 Gamov [133] recalls that "when I was discussing cosmological problems with Einstein, he remarked that the introduction of the cosmological term was the biggest blunder he ever made in his life." In spite of Einstein's regret, the cosmological constant returned at the end of the century to account for the late-time cosmic acceleration.

6.3 The fine tuning problem

In order to realize the cosmic acceleration today, we require that the cosmological constant Λ is of the order of the square of the present Hubble parameter H_0 [see Eq. (6.8)]:

$$\Lambda \approx H_0^2 = (2.1332h \times 10^{-42} \text{ GeV})^2. \quad (6.11)$$

If we interpret this as an energy density, it is equivalent to

$$\rho_\Lambda \approx \frac{\Lambda m_{\text{pl}}^2}{8\pi} \approx 10^{-47} \text{ GeV}^4 \approx 10^{-123} m_{\text{pl}}^4, \quad (6.12)$$

where we have used $h \approx 0.7$ and $m_{\text{pl}} \approx 10^{19} \text{ GeV}$.

Suppose that the energy density (6.12) comes from the vacuum energy $\langle \rho \rangle$ of an empty space. The zero-point energy of some field of mass m with momentum k and frequency ω is given by $E = \omega/2 = \sqrt{k^2 + m^2}/2$ (in the units of $\hbar = c = 1$). Summing over the zero-point energies of this field up to a cut-off scale $k_{\text{max}} (\gg m)$, we obtain the vacuum energy density

$$\rho_{\text{vac}} = \int_0^{k_{\text{max}}} \frac{d^3k}{(2\pi)^3} \frac{1}{2} \sqrt{k^2 + m^2}. \quad (6.13)$$

Since the integral is dominated by the mode with large $k (\gg m)$, we find that

$$\rho_{\text{vac}} = \int_0^{k_{\text{max}}} \frac{4\pi k^2 dk}{(2\pi)^3} \frac{1}{2} \sqrt{k^2 + m^2} \approx \frac{k_{\text{max}}^4}{16\pi^2}. \quad (6.14)$$

General Relativity is believed to be valid up to the Planck scale m_{pl} . Taking the cut-off scale k_{max} to be m_{pl} , the vacuum energy density can be estimated as

$$\rho_{\text{vac}} \simeq 10^{74} \text{ GeV}^4. \quad (6.15)$$

This is about 10^{121} times larger than the observed value (6.12). Note that this situation is not improved much by taking other energy scales appearing in particle physics. For the QCD scale $k_{\text{max}} \approx 0.1 \text{ GeV}$ we have $\rho_{\text{vac}} \approx 10^{-3} \text{ GeV}^4$, which is still much larger than ρ_Λ .

The above problem was present even before the observational discovery of dark energy in 1998. At that time most people believed that the cosmological constant was exactly zero and tried to explain why it was so. The vanishing of a constant usually implies the existence of some symmetry. In supersymmetric theories, for example, the bosonic degree of freedom has its Fermi counterpart that contributes to the zero-point energy with an opposite sign. If supersymmetry is unbroken, there exists an equal number of bosonic and fermionic degrees of freedom such that the total vacuum energy vanishes [134] [see Eq. (6.22) in Section 6.5]. However it is known that supersymmetry is broken at sufficient high energies (around the scale $M_{\text{SUSY}} = 10^3 \text{ GeV}$ if it is relevant to the hierarchy problem of gravitational

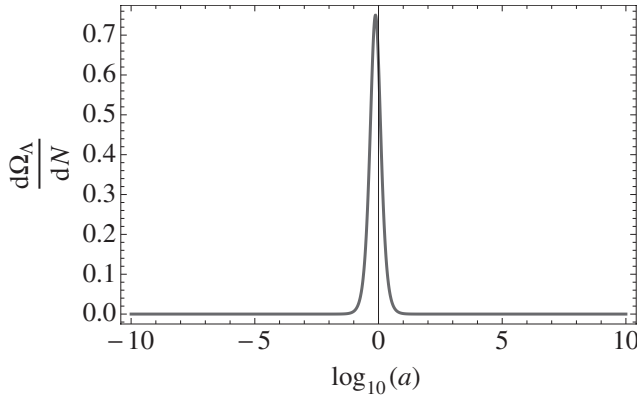


Figure 6.1 Plot of $d\Omega_\Lambda/dN$ versus $\log_{10} a$, assuming flat space with $\Omega_\Lambda^{(0)} = 0.7$. The spike is very close to the present epoch: this is the coincidence problem.

interaction and weak interaction). Hence the vacuum energy is generally non-zero in the world of broken supersymmetry. Nevertheless, it is not impossible to obtain a vanishing Λ or a tiny amount of Λ even if supersymmetry is broken. We shall address this problem in Section 6.5.

6.4 The coincidence problem

The second problem of the cosmological constant as dark energy is that its value is not only at odds with all possible fundamental energy scales and requires therefore fine tuning, but also that this particular value is almost identical to a totally unrelated number, the present matter energy density. In other words, $\Omega_\Lambda^{(0)}$ is doubly unlikely: because it is too small in absolute terms and because its value coincides (to a factor of two or three) with $\Omega_m^{(0)}$, for no obvious reason. The matter density $\rho_m = \rho_m^{(0)}(1+z)^3$ coincides with the cosmological density $\rho_\Lambda^{(0)}$ at

$$z_{\text{coinc}} = \left(\frac{\Omega_\Lambda^{(0)}}{1 - \Omega_\Lambda^{(0)}} \right)^{1/3} - 1, \quad (6.16)$$

which, for $\Omega_\Lambda^{(0)} = 0.7$, amounts to $z_{\text{coinc}} \approx 0.3$. This problem is called the *coincidence problem*.

To illustrate the issue, we plot the evolution of the derivative $d\Omega_\Lambda/dN$ (where $N = \ln a$) in Fig. 6.1. We find that the only epoch in which this quantity is not close to zero is the present.¹ If $\rho_\Lambda^{(0)}/\rho_m^{(0)}$ was just 10 or 100 times smaller, we would not see any accelerated expansion. If it were a few orders of magnitude larger than

¹ We note, however, that the choice of a natural time scale to define the coincidence problem is matter of debate, see e.g. Ref. [135].

unity, the spike would occur at a large redshift and probably we would not call it a coincidence at all.

The coincidence problem is not specific to the cosmological constant. Almost all acceptable dark energy models we will see in the next chapters behave similarly to the cosmological constant and their z_{coinc} also turn out to be very close to zero. Therefore we discuss this problem in terms of a general dark energy density ρ_{DE} .

Barring the case that this coincidence is after all just a coincidence, or that all the observational evidence in favor of acceleration is systematically wrong, cosmologists have proposed several ways out of this problem. The first class of explanations is based on models in which ρ_{DE} responds to the trend of ρ_m and catches up with it irrespective of the initial conditions of ρ_{DE} . In this case Ω_{DE} is non-zero for a considerable duration and this alleviates the coincidence problem. However, the acceleration starts very recently and therefore a coincidence arises again. The problem is in fact that this behavior is based on attractor-like solutions such as the so-called *tracker* models, see Section 7.2.3.

The second class of explanation argues that there is no coincidence and in fact Ω_m and Ω_{DE} have always, or most of the time, been similar. In principle this is not difficult to realize: it is sufficient to postulate two components, one that clusters, the other that does not because of a large sound speed, and to regulate their equations of state so that they are always similar. The main problem here is that either (i) the common equation of state always satisfies the condition for cosmic acceleration and hence it is difficult or impossible to be consistent with many observations such as the growth of large-scale structure, or (ii) the equation of state changes right when it is needed, i.e. today, and therefore another coincidence arises – this time between the epoch of acceleration and the present. Models that belong to this class are for instance the scaling attractors, which will be discussed in Section 8.5.3. A related possibility is to build a model with several epochs of acceleration; it is then just a matter of reasonable chance to be witnessing one. Here again the difficulty is to realize a sufficient period of structure formation.

The third class is the anthropic one. According to it we live in a Universe with $\rho_{\text{DE}} \approx \rho_m$ because this is the highest dark energy density allowed by the requisite of sufficient structure formation and, in general, higher energy vacua are more likely than lower ones. So our Universe is the most likely among the “life-sustaining” universes. We discuss anthropic arguments in Section 6.6.

The fourth class is the “backreaction” argument. The coincidence between ρ_{DE} and ρ_m may appear as a by-product of another more fundamental one, the coincidence between acceleration and structure formation. This can be explained if one causes the other: in particular, if the growth of structures causes acceleration through cumulative non-linear effects. We discuss these models in Section 10.2. A related way out is that actually there is no real acceleration and no dark energy. The acceleration is only an apparent consequence of adopting the wrong background

cosmological model, the FLRW spacetime. If instead we interpret observations with a strongly inhomogeneous model like the Lemaître–Tolman–Bondi void, the acceleration of the recession rate between nearby and distant sources becomes a distance-dependent, but practically always decelerated, Hubble rate. We discuss this interpretation in Section 10.1.

A quick summary of this section is that the coincidence problem is far from solved. It is difficult to imagine a convincing explanation of the nature of dark energy which does not at the same time provide a solution to the coincidence problem. Until then, we can use the coincidence problem, just as the fine-tuning problem, as a guide to select interesting directions of research.

6.5 Supersymmetric models

From here to the end of this section we discuss the status of the cosmological constant in supersymmetric models. This material is more advanced than most of this book and requires knowledge of quantum field theory and supersymmetry. However it is not a necessary prerequisite for the subsequent chapters.

Supersymmetric theoretical models consist of a set of quantum fields having a symmetry between bosons and fermions (see Ref. [136] for introductory review). In particular a supersymmetric model covariant under general coordinate transformations (or a model having local supersymmetry) is called a supergravity model. Supergravity aims to unify the gravitational force with other interactions. When we quantize gravity, the problem of renormalizability is crucial. In renormalizable theories the divergence of integrals which appears in the perturbative expansion for physical processes can be avoided by redefining a finite number of theoretical parameters. Unfortunately supergravity is a non-renormalizable theory and hence it should not be regarded as “a theory of everything.” Superstring theory is the only known theory containing gravity that is renormalizable. Supergravity can be regarded as an effective low-energy theory derived from more fundamental string theory. In the following we shall briefly review supersymmetric theories and then proceed to possible solutions to the cosmological constant problem. The readers who are interested in the details of supergravity and superstring theories may consult the books of Bailin and Love [137] and Green, Schwarz, and Witten [138].

The transformations of supersymmetry are based on quantum operators Q_s which change bosonic states into fermionic ones and vice versa,

$$Q_s|\text{boson}\rangle = |\text{fermion}\rangle, \quad Q_s|\text{fermion}\rangle = |\text{boson}\rangle. \quad (6.17)$$

From this definition the Q_s ’s change the spin of states. The Q_s ’s are identified as spinor operators transforming like tensor operators of spin 1/2. It is known that the Q_s ’s are invariant under coordinate translations. More technically, the Q_s ’s commute with the four generators P^μ of the translation group. We use the

Hermitian generator for $P^\mu = (E, \mathbf{P})$, so that E and $\mathbf{P}$ correspond to the energy operator and the momentum operator, respectively. Then the commutation relation is given by

$$[Q_s, P^\mu] \equiv Q_s P^\mu - P^\mu Q_s = 0. \quad (6.18)$$

The anti-commutator of a generator Q_s with its Hermitian adjoint $Q_s^\dagger$ is a linear combination of energy and momentum operators:

$$\{Q_s, Q_s^\dagger\} \equiv Q_s Q_s^\dagger + Q_s^\dagger Q_s = \alpha E + \beta \mathbf{P}. \quad (6.19)$$

If we sum up all supersymmetry generators, then the $\beta \mathbf{P}$ terms cancel whilst the αE terms add up. This then gives the following dependence

$$\sum_{\text{all } Q_s} \{Q_s, Q_s^\dagger\} = c E. \quad (6.20)$$

Here the proportionality factor c is positive for a physically meaningful theory with energies bounded from below.

The state corresponding to the lowest energy is called a vacuum and is denoted as $|0\rangle$. Supersymmetry is unbroken if the following condition is satisfied:

$$Q_s |0\rangle = 0, \quad \text{and} \quad Q_s^\dagger |0\rangle = 0, \quad \text{for all } Q_s, \quad (6.21)$$

which means that the vacuum is symmetric under the transformation by the Q_s 's. Under the condition (6.21), Eq. (6.20) shows that the vacuum has a vanishing energy

$$E|0\rangle = 0. \quad (6.22)$$

If supersymmetry is broken, the vacuum state is not invariant under all supersymmetry transformations. This means that $Q_s |0\rangle \neq 0$ and $Q_s^\dagger |0\rangle \neq 0$ for some operators Q_s . Hence the energy of the vacuum is not exactly zero.

If the vacuum energy is the only contribution to the cosmological constant Λ , the above discussion shows that $\Lambda \neq 0$ in the world of broken supersymmetry. Note, however, that this result is based on a globally supersymmetric theory without including gravity. If we take into account gravity, any globally supersymmetric theory becomes a locally supersymmetric supergravity theory. In supergravity theory an (effective) cosmological constant is given by an expectation value of the potential V for chiral scalar fields φ^i [137]:

$$V(\varphi, \varphi^*) = e^{\kappa^2 K} [D_i W (K^{ij*}) (D_j W)^* - 3\kappa^2 |W|^2], \quad (6.23)$$

where $\kappa^2 = 8\pi G$, K and W are the so-called Kähler potential and the superpotential, respectively, which are functions of φ^i and φ^{i*} . The quantity K^{ij*} is the

inverse of the following derivative

$$K_{ij^*} \equiv \frac{\partial^2 K}{\partial \varphi^i \partial \varphi^{j^*}}, \quad (6.24)$$

whereas the derivative $D_i W$ is defined by

$$D_i W \equiv \frac{\partial W}{\partial \varphi^i} + \kappa^2 W \frac{\partial K}{\partial \varphi^i}. \quad (6.25)$$

The four-dimensional effective supergravity action is given by

$$S = \int d^4x \sqrt{-g} \left[\frac{1}{2\kappa^2} R - K_{ij^*} \partial_\mu \varphi^i \partial^\mu \varphi^{j^*} - V(\varphi, \varphi^*) \right], \quad (6.26)$$

where the second term represents the kinetic term of chiral scalar fields.

The supersymmetry is preserved under the condition $D_i W = 0$. This gives rise to an anti de Sitter (AdS) minimum for the potential (6.23):

$$V_{\text{AdS}} = -3\kappa^2 e^{\kappa^2 K} |W|^2, \quad (6.27)$$

which is negative. If the conditions $D_i W = 0$ and $W = 0$ hold for all i in lowest order of perturbation theory, it follows that the theory possesses a supersymmetric configuration with $V = 0$ to all orders of perturbation theory [139].

6.5.1 Vanishing cosmological constant under broken supersymmetry

The breaking of the supersymmetry corresponds to the condition $D_i W \neq 0$. In this case it is possible to find scalar field values giving a vanishing potential ($V = 0$), but this is not in general an equilibrium point of the potential V . Nevertheless there is a class of Kähler and superpotentials which give a stationary scalar-field configuration at $V = 0$ (see Ref. [140] for an early work). Consider, for example, the gluino condensation in $E_8 \times E_8$ superstring theory [141]. The reduction of the 10-dimensional action to 4 dimensions gives rise to a so-called modulus field T . This field characterizes the scale of the compactified 6-dimensional manifold. There exists another complex scalar field S of 4-dimensional dilaton/axion fields. The fields T and S are governed by the Kähler potential [141]

$$K(T, S) = -\frac{3}{\kappa^2} \ln(T + T^*) - \frac{1}{\kappa^2} \ln(S + S^*), \quad (6.28)$$

where $(T + T^*)$ and $(S + S^*)$ are positive definite. Here the factor 3 comes from the compactification on a complex manifold with $(10 - 4)/2 = 3$ complex dimensions. The field S couples to the gauge fields, while T does not. An effective superpotential

for S can be obtained by integrating out the gauge fields under the use of the R -invariance [142]:

$$W(S) = M_{\text{pl}}^3 [c_1 + c_2 \exp(-3S/2b)] , \quad (6.29)$$

where c_1 , c_2 , and b are constants.

Substituting Eqs. (6.28) and (6.29) into Eqs. (6.24) and (6.25), it follows that

$$(D_T W) K^{TT*} (D_T W)^* = 3\kappa^2 |W|^2 , \quad (6.30)$$

for the modulus field T . This cancels out the last term in Eq. (6.23), thereby yielding the field potential

$$\begin{aligned} V &= \frac{1}{(T + T^*)^3 (S + S^*)} (D_S W) K^{SS*} (D_S W)^* \\ &= \frac{M_{\text{pl}}^4}{(T + T^*)^3 (S + S^*)} \left| c_1 + c_2 \exp(-3S/2b) \left\{ 1 + \frac{3}{2b} (S + S^*) \right\} \right|^2 . \end{aligned} \quad (6.31)$$

The kinetic term in Eq. (6.26) is given by

$$\mathcal{L}_{\text{kin}} \equiv -K_{ij*} \partial_\mu \varphi^i \partial^\mu \varphi^{j*} \quad (6.32)$$

$$= -\frac{3}{\kappa^2 (T + T^*)^2} \partial_\mu T \partial^\mu T - \frac{1}{\kappa^2 (S + S^*)^2} \partial_\mu S \partial^\mu S . \quad (6.33)$$

For the real scalar fields T and S this reduces to the standard canonical kinetic term, $\mathcal{L}_{\text{kin}} = -(1/2) \partial_\mu \tilde{T} \partial^\mu \tilde{T} - (1/2) \partial_\mu \tilde{S} \partial^\mu \tilde{S}$, by redefining $\tilde{T} \equiv \sqrt{3/2\kappa^2} \ln T$ and $\tilde{S} \equiv \sqrt{1/2\kappa^2} \ln S$.

The potential V in Eq. (6.31) is found to be positive because of the cancellation of the last term in Eq. (6.23). The stationary field configuration with $V = 0$ is realized under the condition

$$D_S W = \frac{\partial W}{\partial S} - \frac{W}{S + S^*} = 0 . \quad (6.34)$$

Note that the derivative,

$$D_T W = \kappa^2 W \frac{\partial K}{\partial T} = -\frac{3W}{T + T^*} , \quad (6.35)$$

does not necessarily vanish. When $D_T W \neq 0$ the supersymmetry is broken with a vanishing potential energy. Hence it is possible to obtain a stationary field configuration with $V = 0$ even if supersymmetry is broken. Note that the field S is fixed at the potential minimum, while the field T is undetermined. The latter appears only for the overall scale of the potential, see Eq. (6.31). Hence such models are called “no-scale” models.

In the above discussion the existence of the term $(-3/\kappa^2) \ln(T + T^*)$ is important to get a positive potential V by canceling out the $-3\kappa^2|W|^2$ term. The appearance of this term is natural in superstring theory after the compactification from 10 to 4 dimensions [137]. In order to get a stationary minimum with $V = 0$ in the world of broken supersymmetry, the fact that the superpotential W depends on the field S is also important. For example, Witten found a model in which W depends on a chiral field C but not on S and T after the compactification of 10-dimensional supergravity on Calabi–Yau manifolds, i.e. $W = W(C)$ [143]. The Kähler potential in this model is given by

$$K = -\frac{3}{\kappa^2} \ln(T + T^* - 2CC^*) - \frac{1}{\kappa^2} \ln(S + S^*). \quad (6.36)$$

In this case we require that $D_S W = 0$ and $D_C W = 0$ to obtain a vanishing potential V . Since W does not depend on S , the condition $D_S W = 0$ leads to $W = 0$ and hence $D_T W = 0$ (see problem 6.1). Then the field configuration with $V = 0$ does not correspond to a broken supersymmetry in this model. This argument shows that the superpotential W is required to have a dependence on S as in the model (6.29) to realize a non-supersymmetric field configuration with $V = 0$.

We caution that the discussion above is based on the lowest-order perturbation theory. It is not guaranteed that this picture is valid to all finite orders of perturbation theory because the non-supersymmetric field configuration is not protected by any symmetry. Moreover some non-perturbative effect can provide a large contribution to the effective cosmological constant. As we will see below, the so-called flux compactification in type IIB string theory allows us to realize a metastable de Sitter vacuum by taking into account a non-perturbative correction to the superpotential (coming from Euclidean D-brane instantons) as well as a number of anti D3-branes in a warped geometry [24]. Hence it is not hopeless to obtain a small value of Λ or a vanishing Λ even in the presence of some non-perturbative corrections. See Refs. [144, 145, 146, 147, 148, 149, 150] for other interesting works that aim to solve the cosmological constant problem.

6.5.2 de Sitter vacua in string theory

Before the observational discovery of dark energy the main interest of cosmologists was to find a mechanism to render the effective cosmological constant zero. After 1998 the interest shifted to finding de Sitter vacua to realize the late-time cosmic acceleration. In the context of superstring or M-theory, the no-go theorem by Gibbons [151] and Maldacena and Nuñez [152] states that when the 6 or 7 extra dimensional space is a time-independent non-singular compact manifold without

boundary a scalar field with a positive potential does not exist at least in the lowest-order action. However the higher-order corrections to the leading-order action or the presence of extended objects like D-branes [153] allow the possibility to realize de Sitter (dS) vacua by invalidating the no-go theorem.

Kachru, Kallosh, Linde, and Trivedi (KKLT) [24] constructed dS solutions in the type II string theory compactified on a Calabi–Yau (CY) manifold in the presence of flux. The construction of the dS vacua in the KKLT scenario consists of two steps. The first step is to freeze all moduli fields in the flux compactification at a supersymmetric anti de Sitter (AdS) vacuum. Then a small number of the anti D3-brane is added in a warped geometry with a throat, so that the AdS minimum is uplifted to yield a de Sitter vacuum with broken supersymmetry.

Let us consider the effective 4-dimensional action derived from the flux compactification of type IIB string theory on the CY manifold. There exist the so-called volume (radial or Kähler) modulus ρ and the complex structure moduli z^α as well as the dilaton field τ . Here the complex structure moduli are associated with the structure of an orientifold in CY theory. In the so-called F -theory compactified on an elliptic CY fourfold, the complex structure moduli z^α of the fourfold can be completely fixed, leaving only the volume modulus ρ . The dilaton field τ , which characterizes the strength of a string coupling g_s , is also fixed under this compactification scheme. For a single volume modulus ρ , the Kähler potential following from the 10-dimensional action is given by

$$K(\rho) = -3 \ln [-i(\rho - \rho^*)] , \quad (6.37)$$

where we use the unit $\kappa^2 = 1$. Note that this is equivalent to the first term on the r.h.s. of Eq. (6.28) by setting $\rho = iT$. The Kähler potential for the dilaton and the complex structure moduli is

$$K(\tau, z^\alpha) = -\ln [-i(\tau - \tau^*)] - \ln \left(-i \int_{\mathcal{M}} \Omega \wedge \Omega^* \right) , \quad (6.38)$$

where Ω is a holomorphic three-form on the CY space $\mathcal{M}$. The expression (6.38) follows from the so-called Weil–Petersson metric, see Ref. [154] for details.

The fluxes generate the following superpotential [155]

$$W_0 = \int_{\mathcal{M}} \Omega \wedge G_{(3)} , \quad (6.39)$$

where $G_{(3)}$ is defined by $G_{(3)} \equiv F_{(3)} - \tau H_{(3)}$ with $F_{(3)}$ and $H_{(3)}$ the Ramond–Ramond (R-R) flux and the Neveu–Schwarz (NS)–NS flux, respectively, on the 3-cycles (non-contractible 3-sphere embedded in the compact manifold) of the

internal CY manifold.² Note that the superpotential (6.39) is independent of ρ . KKLT [24] considered a non-perturbative correction to the superpotential (6.39), which is given by

$$W = W_0 + Ae^{ic\rho}, \quad (6.40)$$

where A and c are constants. This correction is related to the effect of Euclidean D-brane instantons [156].

Since the dilaton τ and the complex structure moduli z^α can be fixed for a suitable choice of flux, we focus on the effective field theory for the volume modulus ρ . We set $\rho = i\sigma$ and take A , c , and W_0 to be real with $W_0 < 0$. For the Kähler potential (6.37) and the superpotential (6.40) the supersymmetric vacuum, $D_\rho W = 0$, corresponds to

$$W_0 = -Ae^{-c\sigma_c} \left(1 + \frac{2}{3}c\sigma_c \right), \quad (6.41)$$

where σ_c is the field value at which the potential V has an AdS minimum.

From Eq. (6.23) the field potential is given by

$$V = \frac{cAe^{-c\sigma}}{2\sigma^2} \left(\frac{1}{3}Ac\sigma e^{-c\sigma} + W_0 + Ae^{-c\sigma} \right). \quad (6.42)$$

The AdS minimum corresponds to the potential energy

$$V_{\text{AdS}} = -\frac{c^2 A^2 e^{-2c\sigma_c}}{6\sigma_c}. \quad (6.43)$$

In Fig. 6.2 we plot the potential (6.42) versus σ for $A = 1$, $c = 0.1$, and $W_0 = -10^{-3}$ as a dotted curve. As estimated by Eq. (6.41) the potential has a negative minimum at $\sigma_c \simeq 88.3$. This minimum corresponds to a supersymmetric AdS vacuum with all moduli fixed (including the dilaton).

KKLT introduced a supersymmetry breaking by adding a few anti D3-brane in a warped background with a throat. The branes as well as fluxes serve as the sources for a warp factor, which allows the possibility to have an exponentially large warping [157]. The Einstein frame metric describing the warped compactification is given by

$$ds_{10}^2 = e^{2A(y)} g_{\mu\nu} dx^\mu dx^\nu + e^{-2A(y)} \tilde{g}_{mn}(y) dy^m dy^n, \quad (6.44)$$

where y characterizes the compact dimension and $\tilde{g}_{mn}$ is the unwarped metric on $\mathcal{M}$. Note that the factor $e^{A(y)}$ can be computed in the region close to a conifold

² The sectors corresponding to periodic boundary conditions and anti-particle boundary conditions are called the Ramond sector and the Neveu–Schwarz sector, respectively. For both left and right moving fermions there are four possible sectors: R-R, NS-NS, NS-R, and R-NS. The spacetime bosonic excitations arise from the RR and NS-NS sectors, whereas the fermions arise from the R-NS and NS-R sectors.

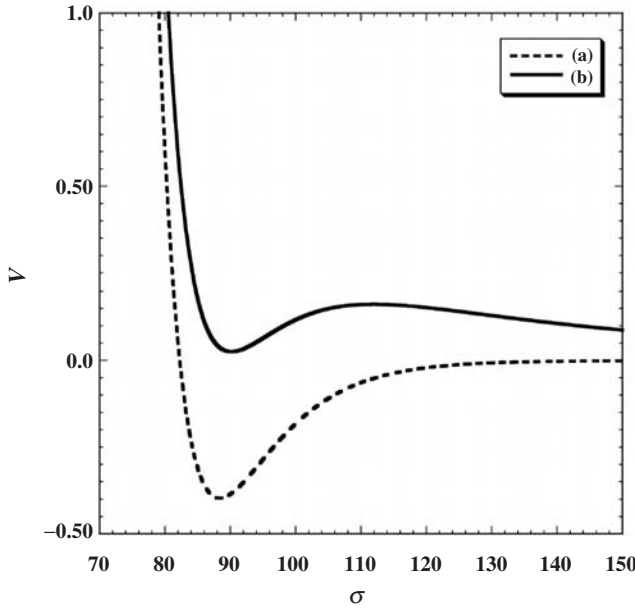


Figure 6.2 The potential of the KKLT model (multiplied by the factor 10^{12}) for the model parameters $A = 1$, $c = 0.1$, $W_0 = -10^{-3}$, and $D = 3.0 \times 10^{-7}$. The dotted curve (a) illustrates the potential (6.42) derived from the Kähler potential (6.37) and the superpotential (6.40). The solid curve (b) shows the potential (6.47) including the D/σ^3 correction to the potential (6.42).

singularity of $\mathcal{M}$. In Ref. [157] it was shown that the warp factor at the tip of the throat is given by

$$e^{A_{\min}} \simeq \exp\left(-\frac{2\pi N}{3g_s M}\right), \quad (6.45)$$

where M and N are the numbers of R-R and NS-NS three-form fluxes, respectively. While the warp factor is of the order of one at generic points in the y -space, the minimum value $e^{A_{\min}}$ can be extremely small for a suitable choice of fluxes.

The addition of the anti-D3 brane does not give rise to additional moduli because the background fluxes generate a potential for the world-volume scalars to be frozen [158]. Meanwhile the anti-D3 brane provides an additional energy to the potential (6.42):

$$\delta V = \frac{2b_0^4 T_3}{g_s^4} \frac{1}{(\text{Im } \rho)^3}, \quad (6.46)$$

where T_3 is the brane tension and b_0 is the warp factor at the location of the anti-D3 brane. The anti-D3 brane energetically prefers to sit at the tip of the throat and

hence $b_0 = e^{A_{\min}}$. The total potential is the sum of Eqs. (6.42) and (6.46),

$$V = \frac{cAe^{-c\sigma}}{2\sigma^2} \left(\frac{1}{3}Ac\sigma e^{-c\sigma} + W_0 + Ae^{-c\sigma} \right) + \frac{D}{\sigma^3}, \quad (6.47)$$

where $D = 2b_0^4 T_3 / g_s^4$.

For suitable choices of D (i.e. by tuning the flux integers M and N) the AdS minimum given in Eq. (6.43) is uplifted by the additional energy (6.46) to give rise to a dS minimum, see the solid curve in Fig. 6.2. There is a local maximum separating the dS minimum from the potential at infinity. As long as D is not chosen to be large, the existence of this local maximum prevents the field σ from evolving toward infinity. If we want to use the dS minimum derived above for the present cosmic acceleration, we require that the potential energy V_{dS} at the minimum is of the order of $V_{\text{dS}} \simeq 10^{-47} \text{ GeV}^4$. Although a fine-tuning of the flux integers M and N is needed, it is in principle possible to obtain a tiny amount of the effective cosmological constant.

The above discussion shows that many de Sitter vacua can be present depending on the flux integers. The question why the vacuum we live in has a very small energy density among many possible vacua has been sometimes answered with anthropic arguments. We discuss some of these ideas in the next section.

6.6 Cosmological constant and the anthropic principle

The anthropic principle is based on the idea that physical theories need to take into account the existence of life on Earth. The expression “anthropic principle” was first used by Carter in 1973 in his contribution to the IAU Krakow symposium [159]. Carter proposed two variants for the anthropic principle.

- (i) Weak anthropic principle (WAP): The WAP states that the spacetime position of life in the Universe is privileged to the extent of being compatible with our existence as observers.
- (ii) Strong anthropic principle (SAP): The SAP states that the Universe and fundamental physical constants must be such as to admit the creation of observers within it at some stage.

Barrow and Tipler [160] applied the WAP to fundamental physical constants by stating that “The observed values of all physical and cosmological quantities are not equally probable but they take on values restricted by the requirement that there exist sites where carbon-based life can evolve and by the requirements that the Universe be old enough for it to have already done so.” According to the definition of Barrow and Tipler, the very low value of the cosmological constant is associated with the WAP rather than SAP unlike the definition of Carter.

Let us discuss the cosmological constant problem in association with the WAP. The formation of large-scale structure needs to be completed before a positive cosmological constant dominates the Universe. Also, a negative cosmological constant leads to the collapse of the Universe if it dominated before the present epoch. Both conditions can be employed to put limits on the anthropically acceptable Λ . In 1987 Weinberg placed the bound on the vacuum energy density to be [161]

$$-10^{-123} m_{\text{pl}}^4 \lesssim \rho_\Lambda \lesssim 3 \times 10^{-121} m_{\text{pl}}^4. \quad (6.48)$$

The upper bound comes from the requirement that the vacuum energy does not dominate over the matter density for redshifts $z \gtrsim \mathcal{O}(1)$, whereas the lower bound comes from the condition that ρ_Λ does not cancel the present cosmological density. Of course the present dark energy density (6.12) lies within Weinberg's bound (6.48).

Even before Weinberg's paper, there had been a number of works about the anthropic principle related to the cosmological constant problem. In the Proceedings of the Nuffield Symposium held in 1982, Linde proposed possible implications of the anthropic principle in inflationary cosmology [162]. This was motivated by the fact that the phase transition in the "new" inflationary scenario [163, 164] can lead to many possible mini-universes isolated from each other. In 1984 Linde suggested an anthropic solution to the cosmological constant problem, which is related to quantum cosmology [165]. The idea is as follows.

Let us assume that quantum creation of the Universe is not suppressed if it is created at the Planck energy density. A scalar field ϕ with a potential energy $V_1(\phi)$ can appear by the quantum creation of the Universe. Also there should exist some classical fields of the type of the anti-symmetric tensor field $A_{\mu\nu\lambda}$. The vacuum energy density $V(\phi, F)$ is the sum of $V_1(\phi)$ and the energy density $V_2(F)$ associated with the field strength $F_{\mu\nu\lambda\sigma} \equiv \partial_{[\mu} A_{\nu\lambda\sigma]}$ of the anti-symmetric tensor field, i.e. $V(\phi, F) = V_1(\phi) + V_2(F)$. The equation of motion for $F^{\mu\nu\lambda\sigma}$ in the absence of sources is given by $\partial_\mu (\sqrt{-g} F^{\mu\nu\lambda\sigma}) = 0$, which has the solution

$$F^{\mu\nu\lambda\sigma} = c \epsilon^{\mu\nu\lambda\sigma}. \quad (6.49)$$

Here c is an arbitrary constant and $\epsilon^{\mu\nu\lambda\sigma}$ is a unit totally anti-symmetric tensor. The square of the 4-form field strength is $F^{\mu\nu\lambda\sigma} F_{\mu\nu\lambda\sigma} = 24c^2$. This gives rise to the energy density

$$V_2(F) = \frac{1}{2 \cdot 4!} F^{\mu\nu\lambda\sigma} F_{\mu\nu\lambda\sigma} = \frac{c^2}{2}, \quad (6.50)$$

which remains constant. The Universe is created most probably in a state with $V(\phi, F) \simeq m_{\text{pl}}^4$. After the field ϕ rolls down to its potential minimum at $\phi = \phi_0$,

the vacuum energy takes a completely different value

$$\Lambda = V_1(\phi_0) + V_2(F). \quad (6.51)$$

Note that $V_1(\phi)$ can take any initial value such that $V_1(\phi) + V_2(F) \simeq m_{\text{pl}}^4$. Then the final value of the cosmological constant Λ appears with approximately the same probability. This may provide a possible anthropic solution to the cosmological constant problem.

In 1987–1988 Brown and Teitelboim tried to find a mechanism to make the effective cosmological constant variable by using the 4-form field $F^{\mu\nu\lambda\sigma}$ [166, 167] (see also Ref. [168]). This anticipated the work of Bousso and Polchinski in 2000 [169] as well as the landscape of string theory [25]. In string theory there are “electric charges” (membranes) sourcing the 4-form field dual to “magnetic charges” (5-branes). The constant c appearing in Eq. (6.49) can be quantized in integer multiples of the membrane charge q :

$$c = nq. \quad (6.52)$$

Then the energy density (6.50) for the 4-form field yields

$$V_2(F) = \frac{n^2 q^2}{2}. \quad (6.53)$$

Let us consider a bare cosmological constant Λ_b in addition to the flux energy density (6.53). Then the effective cosmological constant is given by

$$\Lambda = \Lambda_b + \frac{n^2 q^2}{2}. \quad (6.54)$$

This is similar to Eq. (6.51) apart from the fact that the constant c is discrete because of the quantization.

As in the KKLT model, let us consider the AdS minimum ($\Lambda_b < 0$) in the presence of the flux energy density $n^2 q^2/2$. The membranes can spontaneously appear by a quantum tunneling effect (which is similar to the appearance of an electron and a positron from the tunneling effect out of the vacuum). The field strength of the 4-form field is slowly discharged by such a Schwinger pair creation of field sources [$nq \rightarrow (n-1)q$]. This leads to the decrease of the vacuum energy of the 4-form field by a discrete amount

$$\delta\Lambda = [n^2 q^2 - (n-1)^2 q^2]/2 = (n-1/2)q^2. \quad (6.55)$$

In order to get the smallest value of Λ in Eq. (6.54) we need to take the flux integer n_0 nearest to $\sqrt{2|\Lambda_b|}/q$. Then the step size near $\Lambda = 0$ is $(n_0 - 1/2)q^2 \simeq q\sqrt{|\Lambda_b|}$. From the requirement that this is of the order $\rho_\Lambda \simeq 10^{-123}$ or smaller in Planck

units ($m_{\text{pl}} = 1$), this relation provides the constraint

$$q \lesssim 10^{-123} |\Lambda_b|^{-1/2}. \quad (6.56)$$

Since the bare cosmological constant cannot be taken so small, Eq. (6.56) shows that the membrane charge q is constrained to be unnaturally small. The anthropic principle may state that such a small charge is selected for our existence, but the problem is that there is no consistent mechanism to obtain such a tiny charge in string theory.

Bousso and Polchinski [169] tackled this problem by considering J (> 1) 4-form fields together with J membrane species with charges $q_1, q_2, \dots, q_J$. The configuration with J of the order of 100 naturally appears in the context of string theory. Using the quantized field strength, $F_i^{\mu\nu\lambda\sigma} = n_i q_i \epsilon^{\mu\nu\lambda\sigma}$, the effective cosmological constant is given by

$$\Lambda = \Lambda_b + \frac{1}{2} \sum_{i=1}^J n_i^2 q_i^2, \quad (6.57)$$

which is the generalization of Eq. (6.54) to multiple 4-form fields. It is possible to obtain the value $\Lambda \simeq 10^{-123}$ if there exists a set of integers n_i such that

$$2|\Lambda_b| < \sum_{i=1}^J n_i^2 q_i^2 < 2(|\Lambda_b| + \Delta\Lambda), \quad (6.58)$$

where $\Delta\Lambda \simeq 10^{-123}$. Let us consider a J -dimensional grid with axes corresponding to $n_i q_i$. The displacement of the 4-form field is given by discrete grid points with integers n_i . The region described by Eq. (6.58) corresponds to a thin-shell with radius $r = \sqrt{2|\Lambda_b|}$ and width $\Delta r = \Delta\Lambda / \sqrt{2|\Lambda_b|}$. When $J = 2$, for example, the thin-shell has an area with $V_S = 2\pi r \Delta r = 2\pi \Delta\Lambda$. For general J the thin-shell has a volume

$$V_S = \Omega_{J-1} r^{J-1} \Delta r = \Omega_{J-1} |2\Lambda_b|^{J/2-1} \Delta\Lambda, \quad (6.59)$$

where $\Omega_{J-1} = 2\pi^{J/2} / \Gamma(J/2)$ is the area of a unit $J - 1$ dimensional sphere ($\Gamma(x)$ is the Gamma function). The volume of a grid cell is given by

$$V_C = \prod_{i=1}^J q_i. \quad (6.60)$$

There is at least one value of Λ if $V_C < V_S$, i.e.

$$\prod_{i=1}^J q_i < \frac{2\pi^{J/2}}{\Gamma(J/2)} |2\Lambda_b|^{J/2-1} \Delta\Lambda. \quad (6.61)$$

When $J = 100$, $|\Lambda_b| = 1$ and $\Delta V = 10^{-123}$ with equal charges ($q_i = q$, for $i = 1, 2, \dots, J$), we find that the condition (6.61) can be satisfied for $q < 0.035$. Since the charge $\sqrt{q}$ has the dimension of mass from Eq. (6.57), this condition translates into $\sqrt{q} < 0.19$ in the unit of the Planck mass. Hence the natural values of the charge q are allowed by considering multiple 4-form fields.

In string theory, 6-dimensional compactified manifolds (such as Calabi–Yau manifolds) have hundreds of different 3-cycles. A 5-brane wrapping a 3-cycle can be viewed as a 2-brane (membrane) to a macroscopic observer. The 5-brane can wrap any of these 3-cycles, which gives rise to hundreds of different membranes in four-dimensional spacetime. The charges q_i are determined by the 5-brane charge, the volume of compactified manifolds $\mathcal{M}$, and the 3-cycle volume. Generally the charges q_i are slightly smaller than unity, which is consistent with the discussion above. Since the volumes of the 3-cycles are in general different from each other, the q_i s also differ from each other.

The number of vacua appearing in string theory can be enormously large. If we consider 500 3-cycles with each cycle wrapped by up to 10 fluxes, then there will be 10^{500} vacua. The fact that there is a large number of possible de Sitter vacua in string theory has led to the so-called string landscape [25]. This landscape includes so many possible configurations of local minima, among which our Universe corresponds to at least one of them. Then, the argument goes on, it should be possible to find a vacuum with a very small energy density among 10^{500} vacua. In this way the landscape of string theory can be used as a concrete implementation of the anthropic principle.

Each vacuum with flux configuration $(n_1, \dots, n_i, \dots, n_J)$ is stable at least at the classical level. At the quantum level the spontaneous creation of a membrane can change the number of fluxes. As we have discussed in Eq. (6.55), the cosmological constant will jump by the amount $(n_i - 1/2)q_i^2$ after the spontaneous creation of membranes. Since this tunneling process is generally exponentially suppressed, the discrete vacua have sufficiently long lifetimes. Note that the KKLT model also has a large amount of such metastable vacua. In Ref. [24] it was shown that the lifetimes of the KKLT vacua are much larger than the cosmological time scale given in Eq. (2.36).

The picture of string landscape has changed our way of thinking how string theory makes other predictions. Each vacuum in the landscape has different matter and coupling constants. The standard model is not predicted uniquely in this picture. We need to perform a statistical prediction for our vacuum relative to the abundance in the landscape. Some people have studied landscape statistics by considering the relative abundance of long-lived low-energy vacua [170, 171]. In the context of eternal inflation (producing different vacua in widely separate regions of the Universe) some vacua are selected preferably compared to others, whose properties

enter in the statistical predictions [172, 173, 174]. These statistical approaches are still under study, but it will be interesting to pursue the possibility to obtain high probabilities for the appearance of low-energy vacua.

Ultimately all these approaches make use of Weinberg's bound or something similar to select values of Λ compatible with life or observers or simply sufficient complexity. One has to notice however that Weinberg's argument assumes a standard scenario of structure formation. A much faster growth of perturbations could be compatible with a much larger Λ since then structures have the time to form before the Λ -domination. This reflects a general problem with anthropic arguments, namely that often they are applied to a single parameter while fixing all the others (instead of marginalizing over them, using the Bayesian language of Chapter 13). A parameter value that is "ruled out" in one case could be acceptable if something else is changed at the same time. Since "intelligent life" is a very complex phenomenon, it is likely to be related to a vast number of parameters and it is hard to see how one could consider all possibilities.

6.7 The decoupling of the cosmological constant from gravity

Whether or not the vacuum energy we have evaluated in Eq. (6.14) really contributes to dark energy is still a debatable problem. Usually this energy can be eliminated by the normal ordering prescriptions in quantum field theory or it can be normalized to any value. If we really try to link the vacuum energy with dark energy, the theory should be enlarged to include gravity.

Usually the gravitational field equations are obtained from the variation of the action

$$S = \int d^4x \sqrt{-g} [\mathcal{L}_{\text{grav}}(g_{\mu\nu}) + \mathcal{L}_m(g_{\mu\nu}, \Psi_i)] , \quad (6.62)$$

by varying the metric $g_{\mu\nu}$, where $\mathcal{L}_{\text{grav}}(g_{\mu\nu})$ is a gravitational Lagrangian which depends on the metric and $\mathcal{L}_m(g_{\mu\nu}, \Psi_i)$ is a matter Lagrangian dependent on both the metric $g_{\mu\nu}$ and the matter fields Ψ_i . The resulting gravitational field equations are of the form $E_{\mu\nu} = \kappa^2 T_{\mu\nu}$, where $E_{\mu\nu}$ is some geometric quantity (in Einstein theory $E_{\mu\nu} = G_{\mu\nu}$) and $T_{\mu\nu}$ is the energy-momentum tensor. The theory is not invariant under the shift of the energy-momentum tensor: $T_{\mu\nu} \rightarrow T_{\mu\nu} + \Lambda g_{\mu\nu}$. This gives rise to the additional contribution from Λ to the Friedmann equation (2.17). Note that in the absence of gravity the equation of motion for matter is invariant under such a transformation. In this case the value of Λ is undetermined. This means that gravity breaks the symmetry which is present in the matter sector without gravity. The root of the cosmological constant problem is intimately related

to the fact that the conventional gravitational theory derived by the variation of the metric $g_{\mu\nu}$ does not allow a symmetry under the shift $T_{\mu\nu} \rightarrow T_{\mu\nu} + \Lambda g_{\mu\nu}$ [175].

If we obtain a theory in which such a symmetry is respected, then the vacuum energy can be decoupled from gravity. This means that the bulk cosmological constant is gravitationally inactive, which allows the possibility to explain why the bulk gravitational constant is irrelevant and may be taken to be zero. A realization of this possibility called *degravitation* of the vacuum has been discussed by a number of authors. In one of them, known as *cascading gravity* [176, 177], gravity is shut off for large-scale density perturbations, whereas another possibility is to introduce an incompressible *gravitational Aether* fluid to degravitate the vacuum [178]. In a series of works Padmanabhan [175, 179, 180] has shown how to construct a theory which has a symmetry under the shift $T_{\mu\nu} \rightarrow T_{\mu\nu} + \Lambda g_{\mu\nu}$. In the following we shall focus on this latter possibility.

Instead of taking the metric $g_{\mu\nu}$ to vary the action, we take normalized vector fields $n_a(x)$ in the spacetime with a norm fixed at every event, i.e. $n_a n^a = \epsilon(x)$ where $\epsilon(x)$ is a fixed function. This is motivated by the analogies with the ordinary solids in elasticity – the macroscopic elastic dynamics is described by a displacement vector field which occurs in the equation $x^\mu \rightarrow x^\mu + \xi^\mu(x)$. As in the displacement vector $\xi^\mu(x)$, the vector $n_a(x)$ captures the spacetime macroscopic description for gravity. The general action to describe the effective description of a D -dimensional spacetime is [179, 180]

$$S = S_{\text{grav}} + S_m, \quad (6.63)$$

where

$$\begin{aligned} S_{\text{grav}} &= -4 \int_{\mathcal{V}} d^D x \sqrt{-g} P_{ab}{}^{cd} \nabla_c n^a \nabla_d n^b, \\ S_m &= \int_{\mathcal{V}} d^D x \sqrt{-g} T_{ab} n^a n^b, \end{aligned} \quad (6.64)$$

where $\mathcal{V}$ is a spacetime volume, T_{ab} is an energy-momentum tensor of matter, and $P_{ab}{}^{cd}$ is a geometric tensor specified later.

We vary the action (6.63) with respect to n_a and derive the equations of motion according to $\delta S / \delta n_a = 0$. Such a procedure is possible if (i) the tensor P^{abcd} has algebraic symmetries (such as $P^{abcd} = -P^{abdc}$) similar to the Riemann tensor R^{abcd} and (ii) the tensor P^{abcd} satisfies the relation $\nabla_a P^{abcd} = 0$ [179]. An explicit form of P^{abcd} should be determined by the macroscopic limit of some microscopic theory, but such a theory is absent so far. Alternatively, we expand P^{abcd} in powers of derivatives of the metric

$$P^{abcd}(g_{ij}, R_{ijkl}) = c_0 P^{(0)abcd}(g_{ij}) + c_1 P^{(1)abcd}(g_{ij}, R_{ijkl}) + \dots, \quad (6.65)$$

where c_0 and c_1 are constants. The zero-th-order term depends on the metric g_{ij} only. The m -th-order term is dependent on the m -th-order curvature tensor as well as g_{ij} . The tensor P^{abcd} can be uniquely determined as

$$P_{ab}^{(m)cd} \propto \delta_{abb_3\dots b_{2m}}^{cd a_3\dots a_{2m}} R_{a_3 a_3}^{b_3 b_4} \dots R_{a_{2m-1} a_{2m}}^{b_{2m-1} b_{2m}}, \quad (6.66)$$

where $\delta_{abb_3\dots b_{2m}}^{cd a_3\dots a_{2m}}$ is an alternating tensor (i.e. a completely anti-symmetric tensor). The zero-th-order term is

$$P_{cd}^{(0)ab} = \frac{1}{16\pi} \frac{1}{2} \delta_{b_1 b_2}^{a_1 a_2} = \frac{1}{32\pi} (\delta_c^a \delta_d^b - \delta_d^a \delta_c^b), \quad (6.67)$$

where the coefficient is chosen to lead to Einstein theory (as we will see below). The first-order term can give rise to the Gauss–Bonnet curvature corrections [see Eq. (9.163) for the definition of the Gauss–Bonnet term].

Let us vary the action (6.63) with respect to the normalized vector field n^a under the constraint $\delta(n^a n_a) = 0$. Imposing the constraint $\delta(n^a n_a) = 0$ allows the possibility to add the integral of the form $\int d^D x \sqrt{-g} \lambda(x) g_{ab} n^a n^b$ to the action (6.63). Then the variation of the modified action gives

$$\delta S = -2 \int_{\mathcal{V}} d^D x \sqrt{-g} [4 P_{ab}{}^{cd} \nabla_c n^a (\nabla_d \delta n^b) - T_{ab} n^a \delta n^b - \lambda g_{ab} n^a \delta n^b], \quad (6.68)$$

where we have used $P_{ab}{}^{cd} = P_{ba}{}^{dc}$ and $T_{ab} = T_{ba}$. Integrating this equation by parts under the use of the condition $\nabla_d P_{ab}{}^{cd} = 0$, we find

$$\begin{aligned} \delta S = & 2 \int_{\mathcal{V}} d^D x \sqrt{-g} [4 P_{ab}{}^{cd} (\nabla_d \nabla_c n^a) + (T_{ab} + \lambda g_{ab}) n^a] \delta n^b \\ & - 8 \int_{\partial \mathcal{V}} d^{D-1} x \sqrt{h} [\ell_d P_{ab}{}^{cd} (\nabla_c n^a)] \delta n^b, \end{aligned} \quad (6.69)$$

where ℓ_d is a vector normal to the spacetime boundary (horizon) $\partial \mathcal{V}$, and h is the determinant of an intrinsic metric on $\partial \mathcal{V}$. For the validity of the variational principle we require that the vector variation δn^a vanishes at the boundary, so that the last term in Eq. (6.69) disappears. Using the anti-symmetric relation, $P_{ab}{}^{cd} = -P_{ab}{}^{dc}$, the variational principle $\delta S = 0$ gives

$$2 P_{ab}{}^{cd} (\nabla_c \nabla_d - \nabla_d \nabla_c) n^a - (T_{ab} + \lambda g_{ab}) n^a = 0. \quad (6.70)$$

Since the Riemann tensor $R_{cde}{}^a$ satisfies $(\nabla_c \nabla_d - \nabla_d \nabla_c) n^a = -R_{cde}{}^a n^e$, Eq. (6.70) reduces to

$$2 P_{ab}{}^{cd} R_{cde}{}^a n^e + (T_{ab} + \lambda g_{ab}) n^a = 0. \quad (6.71)$$

Multiplying $g^{eb}g_{ae}$ for Eq. (6.71) and changing the indices suitably, we obtain

$$\left(2P_b^{ijk}R^a_{ijk} - T_b^a - \lambda\delta_b^a\right)n_a = 0. \quad (6.72)$$

Demanding that the condition (6.72) holds for arbitrary vector fields n_a , it follows that

$$2P_b^{ijk}R^a_{ijk} - T_b^a - \lambda\delta_b^a = 0. \quad (6.73)$$

Plugging the zero-th-order term (6.67) into Eq. (6.73), we find

$$\frac{1}{8\pi G}R_b^a - T_b^a - \lambda\delta_b^a = 0. \quad (6.74)$$

Using the Einstein tensor $G_b^a = R_b^a - (1/2)\delta_b^a R$, Eq. (6.74) can be written as $G_b^a - 8\pi GT_b^a = F(g)\delta_b^a$ with $F(g) = -8\pi G\lambda - R/2$. Because of the constraint $\nabla_a(G_b^a - 8\pi GT_b^a) = 0$ it follows that $\partial_b F = \partial_b(-8\pi G\lambda - R/2) = 0$. Hence F is an integration constant, say Λ . This gives the relation $\Lambda = -8\pi G\lambda - R/2$. The resulting field equation is given by

$$G_b^a = 8\pi GT_b^a + \Lambda\delta_b^a, \quad (6.75)$$

which corresponds to the Einstein equations with the cosmological constant Λ . Interestingly the cosmological constant appears as an integration constant. Note that the first-order term (6.65) gives rise to the Gauss–Bonnet curvature correction in addition to the field equation (6.75).

If we make the shift $T_{ab} \rightarrow T_{ab} + \rho g_{ab}$ for the energy-momentum tensor, the matter action S_m in Eq. (6.64) is transformed as

$$\int_{\mathcal{V}} d^D x \sqrt{-g} T_{ab} n^a n^b \rightarrow \int_{\mathcal{V}} d^D x \sqrt{-g} T_{ab} n^a n^b + \int_{\mathcal{V}} d^D x \sqrt{-g} \epsilon \rho. \quad (6.76)$$

The quantity ϵ is not varied when we vary n_a , which means that the theory is invariant under the shift $T_{ab} \rightarrow T_{ab} + \rho g_{ab}$. Hence it is possible to gauge away the cosmological constant from gravity according to the variational principle with respect to the normalized vector n_a . Naturally, we still have the problem to explain why the observed dark energy has a very small energy scale. If the cosmological constant is completely decoupled from gravity, we need to find out alternative models of dark energy consistent with observations.

In the following chapters we shall consider alternative models of dark energy under the assumption that the cosmological constant problem is solved in such a way that it vanishes completely.

6.8 Problems

- 6.1** For the Kähler potential (6.36) and the superpotential $W(C)$ that depends on the field C only, derive the field potential V . Show that the field configuration with $V = 0$ corresponds to the supersymmetric state.
- 6.2** Derive the potential (6.42) from the Kähler potential (6.37) and the superpotential (6.40). Show that the potential (6.42) has a minimum satisfying the condition (6.41).

Dark energy as a modified form of matter I: Quintessence

If the cosmological constant problem is solved in a way that Λ completely vanishes, we need to find alternative models of dark energy. As we already mentioned in the Introduction, there are basically two approaches for the construction of dark energy models. The first approach is based on “modified matter models” in which the energy-momentum tensor $T_{\mu\nu}$ on the r.h.s. of the Einstein equations contains an exotic matter source with a negative pressure. The second approach is based on “modified gravity models” in which the Einstein tensor $G_{\mu\nu}$ on the l.h.s. of the Einstein equations is modified.

It is however important to realize that within General Relativity this division is mostly a practical way to classify the variety of dark energy models but does not carry a fundamental meaning. One can always write down Einstein’s equations in the standard form $G_{\mu\nu} = 8\pi G T_{\mu\nu}$ by absorbing in $T_{\mu\nu}$ all the gravity modifications that one conventionally puts on the l.h.s.. In other words, one can define a covariantly conserved energy-momentum tensor that equals the Einstein tensor. There is no way, within General Relativity, i.e. by using only gravitational interactions, to distinguish modified matter from modified gravity. At first-order in perturbation theory, for instance, one can define the equation of state and the sound speed of a dark energy field so that it reproduces any modified gravity model [181]. Of course, from the viewpoint of quantum field theory the situation is different and the field content of the two classes are in general different and in principle distinguishable.

In this chapter we study the so-called *quintessence* model as one of the representative modified matter models. We will discuss other modified matter models in Chapter 8 and modified gravity models in Chapter 9.

Caldwell *et al.* [35] named quintessence a canonical scalar field ϕ with a potential $V(\phi)$ responsible for the late-time cosmic acceleration.¹ Unlike the cosmological

¹ According to ancient Greek science, the quintessence (from the Latin “fifth element”) denotes a fifth cosmic element after earth, fire, water, and air. See the Introduction.

constant, the equation of state of quintessence dynamically changes with time. The cosmological dynamics for quintessence in the presence of matter and radiation has a long history – already in the 1980s the cosmological consequences for such a system had been discussed by a number of authors [26, 28, 29, 30]. The cosmological evolution can be easily understood by a dynamical system approach, as we will see in this chapter.

The energy density of quintessence does not need to be very small with respect to radiation or matter in the early Universe unlike the cosmological constant scenario. The existence of the so-called “tracker” field [39] is important to alleviate the coincidence problem of dark energy. The tracker fields correspond to attractor-like solutions in which the field energy density tracks the background fluid density for a wide range of initial conditions. We shall discuss conditions under which such tracking behavior occurs. The constraint on the quintessence energy density in the early cosmological epoch will be also discussed.

There have been many attempts to construct particle physics models of quintessence. In so doing one needs to find field potentials flat enough to lead to the slow-roll inflation today with an energy scale $\rho_{\text{DE}} \simeq 10^{-123} m_{\text{pl}}^4$ and a mass scale $m_\phi \lesssim 10^{-33}$ eV. Although this is an obstacle for the model building, it is not impossible to construct viable models of quintessence in the framework of particle physics. We shall discuss this issue in considerable detail.

7.1 Quintessence

Many scalar fields are present in particles physics – including string theory and supergravity. We use the term “quintessence” to denote a canonical scalar field ϕ with a potential $V(\phi)$ that interacts with all the other components only through standard gravity. The quintessence model is therefore described by the action

$$S = \int d^4x \sqrt{-g} \left[\frac{1}{2\kappa^2} R + \mathcal{L}_\phi \right] + S_M, \quad \mathcal{L}_\phi = -\frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - V(\phi), \quad (7.1)$$

where $\kappa^2 = 8\pi G$ and R is the Ricci scalar. Note that we have taken into account the matter action S_M .

We consider a perfect fluid with the energy density ρ_M , the pressure P_M , and the equation of state $w_M = P_M/\rho_M$. Here the subscript “ M ” is used for a general perfect fluid (including the case of a total fluid) without specifying non-relativistic matter or radiation. Later we shall use the subscript “ m ” to specify non-relativistic matter. The fluid satisfies the continuity equation (2.20), i.e.

$$\dot{\rho}_M + 3H(\rho_M + P_M) = 0. \quad (7.2)$$

The energy-momentum tensor of quintessence is [73]

$$T_{\mu\nu}^{(\phi)} = -\frac{2}{\sqrt{-g}} \frac{\delta(\sqrt{-g}\mathcal{L}_\phi)}{\delta g^{\mu\nu}} \quad (7.3)$$

$$= \partial_\mu \phi \partial_\nu \phi - g_{\mu\nu} \left[\frac{1}{2} g^{\alpha\beta} \partial_\alpha \phi \partial_\beta \phi + V(\phi) \right]. \quad (7.4)$$

In a FLRW background, the energy density ρ_ϕ and the pressure P_ϕ of the field are

$$\rho_\phi = -T_0^{0(\phi)} = \frac{1}{2} \dot{\phi}^2 + V(\phi), \quad P_\phi = \frac{1}{3} T_i^i{}^{(\phi)} = \frac{1}{2} \dot{\phi}^2 - V(\phi), \quad (7.5)$$

which give the equation of state

$$w_\phi \equiv \frac{P_\phi}{\rho_\phi} = \frac{\dot{\phi}^2 - 2V(\phi)}{\dot{\phi}^2 + 2V(\phi)}. \quad (7.6)$$

In the flat Universe ($K = 0$) the following equations of motion follow from Eqs. (2.17) and (2.18):

$$H^2 = \frac{\kappa^2}{3} \left[\frac{1}{2} \dot{\phi}^2 + V(\phi) + \rho_M \right], \quad (7.7)$$

$$\dot{H} = -\frac{\kappa^2}{2} (\dot{\phi}^2 + \rho_M + P_M), \quad (7.8)$$

where $\kappa^2 = 8\pi G$. The variation of the action (7.1) with respect to ϕ gives

$$\ddot{\phi} + 3H\dot{\phi} + V_{,\phi} = 0, \quad (7.9)$$

where $V_{,\phi} \equiv dV/d\phi$. The Klein–Gordon equation (7.9) can be also derived by using the continuity equation $\dot{\rho}_\phi + 3H(\rho_\phi + P_\phi) = 0$ or by combining Eqs. (7.2), (7.7), and (7.8).

During radiation- or matter-dominated epochs, the energy density ρ_M of the fluid dominates over that of quintessence, i.e. $\rho_M \gg \rho_\phi$. We require that ρ_ϕ tracks ρ_M so that the dark energy density emerges at late times. Whether this tracking behavior occurs or not depends on the form of the potential $V(\phi)$. If the potential is steep so that the condition $\dot{\phi}^2/2 \gg V(\phi)$ is always satisfied, the field equation of state is given by $w_\phi \simeq 1$ from Eq. (7.6). In this case the energy density of the field evolves as $\rho_\phi \propto a^{-6}$, which decreases much faster than the background fluid density.

We require the condition $w_\phi < -1/3$ to realize the late-time cosmic acceleration, which translates into the condition $\dot{\phi}^2 < V(\phi)$. Hence the scalar potential needs to be shallow enough for the field to evolve slowly along the potential. This situation is similar to that in inflationary cosmology and it is convenient to introduce the

following slow-roll parameters [94]

$$\epsilon_s \equiv \frac{1}{2\kappa^2} \left(\frac{V_{,\phi}}{V} \right)^2, \quad \eta_s \equiv \frac{V_{,\phi\phi}}{\kappa^2 V}. \quad (7.10)$$

If the conditions $\epsilon_s \ll 1$ and $|\eta_s| \ll 1$ are satisfied, the evolution of the field is sufficiently slow so that $\dot{\phi}^2 \ll V(\phi)$ and $|\ddot{\phi}| \ll |3H\dot{\phi}|$ in Eqs. (7.7) and (7.9).

From Eq. (7.9) the deviation of w_ϕ from -1 is given by

$$1 + w_\phi = \frac{V_{,\phi}^2}{9H^2(\xi_s + 1)^2 \rho_\phi}, \quad (7.11)$$

where $\xi_s \equiv \ddot{\phi}/(3H\dot{\phi})$. This shows that w_ϕ is always larger than -1 for a positive potential. In the slow-roll limit, $|\xi_s| \ll 1$ and $\dot{\phi}^2/2 \ll V(\phi)$, we obtain $1 + w_\phi \simeq 2\epsilon_s/3$ by neglecting the matter fluid in Eq. (7.7), i.e. $3H^2 \simeq \kappa^2 V(\phi)$. The deviation of w_ϕ from -1 is characterized by the slow-roll parameter ϵ_s .

So far many quintessence potentials have been proposed. Crudely speaking they have been classified into (i) “freezing models” and (ii) “thawing” models [182]. In class (i) the field was rolling along the potential in the past, but the movement gradually slows down after the system enters the phase of cosmic acceleration. The representative potentials that belong to this class are

(i) Freezing models

- $V(\phi) = M^{4+n} \phi^{-n} \quad (n > 0),$
- $V(\phi) = M^{4+n} \phi^{-n} \exp(\alpha \phi^2/m_{\text{pl}}^2).$

The first potential does not possess a minimum and hence the field rolls down the potential toward infinity [30, 38]. This appears, for example, in the fermion condensate model as a dynamical supersymmetry breaking [48]. The second potential has a minimum at which the field is eventually trapped (corresponding to $w_\phi = -1$). This potential can be constructed in the framework of supergravity [50].

In class (ii) the field (with mass m_ϕ) has been frozen by Hubble friction (i.e. the term $H\dot{\phi}$) until recently and then it begins to evolve once H drops below m_ϕ . The equation of state of dark energy is $w_\phi \simeq -1$ at early times, which is followed by the growth of w_ϕ . The representative potentials that belong to this class are

(ii) Thawing models

- $V(\phi) = V_0 + M^{4-n} \phi^n \quad (n > 0),$
- $V(\phi) = M^4 \cos^2(\phi/f).$

The first potential is similar to the one of chaotic inflation ($n = 2, 4$) used in the early Universe (with $V_0 = 0$) [183], while the mass scale M is very different. Note

that the model with $n = 1$ was originally proposed in Ref. [184] to replace the cosmological constant by a slowly varying field and was revised in Ref. [185] in connection with the possibility to allow for negative values of $V(\phi)$. The Universe will collapse in the future if the system enters the region with $V(\phi) < 0$. The second potential appears as a potential for the Pseudo-Nambu–Goldstone Boson (PNGB). This was introduced by Frieman *et al.* [47] in response to the first tentative suggestions that the Universe may be dominated by the cosmological constant. In this model the field is nearly frozen at the potential maximum during the period in which the field mass m_ϕ is smaller than H , but it begins to roll down around the present ($m_\phi \simeq H_0$).

7.2 Dynamical system approach

In order to study cosmological dynamics in the presence of a scalar field and a background fluid, it is convenient to introduce the following dimensionless variables:

$$x_1 \equiv \frac{\kappa \dot{\phi}}{\sqrt{6}H}, \quad x_2 \equiv \frac{\kappa \sqrt{V}}{\sqrt{3}H}. \quad (7.12)$$

Then Eq. (7.7) can be written as

$$\Omega_M \equiv \frac{\kappa^2 \rho_M}{3H^2} = 1 - x_1^2 - x_2^2. \quad (7.13)$$

We also define the energy fraction of dark energy

$$\Omega_\phi \equiv \frac{\kappa^2 \rho_\phi}{3H^2} = x_1^2 + x_2^2, \quad (7.14)$$

which satisfies the relation $\Omega_M + \Omega_\phi = 1$. From Eq. (7.8) we obtain

$$\frac{\dot{H}}{H^2} = -3x_1^2 - \frac{3}{2}(1 + w_M)(1 - x_1^2 - x_2^2), \quad (7.15)$$

in which case the effective equation of state (4.96) is given by

$$w_{\text{eff}} = w_M + (1 - w_M)x_1^2 - (1 + w_M)x_2^2. \quad (7.16)$$

The equation of state of dark energy (7.6) can be expressed as

$$w_\phi = \frac{x_1^2 - x_2^2}{x_1^2 + x_2^2}. \quad (7.17)$$

Differentiating the variables x_1 and x_2 with respect to the number of e-foldings $N = \ln a$ together with the use of Eqs. (7.9) and (7.15), we obtain the following

equations

$$\frac{dx_1}{dN} = -3x_1 + \frac{\sqrt{6}}{2}\lambda x_2^2 + \frac{3}{2}x_1[(1 - w_M)x_1^2 + (1 + w_M)(1 - x_2^2)] , \quad (7.18)$$

$$\frac{dx_2}{dN} = -\frac{\sqrt{6}}{2}\lambda x_1 x_2 + \frac{3}{2}x_2[(1 - w_M)x_1^2 + (1 + w_M)(1 - x_2^2)] , \quad (7.19)$$

where

$$\lambda \equiv -\frac{V_{,\phi}}{\kappa V} . \quad (7.20)$$

The quantity λ characterizes the slope of the field potential, which obeys the following equation

$$\frac{d\lambda}{dN} = -\sqrt{6}\lambda^2(\Gamma - 1)x_1 , \quad (7.21)$$

where

$$\Gamma \equiv \frac{V V_{,\phi\phi}}{V_{,\phi}^2} . \quad (7.22)$$

If λ is constant, the integration of Eq. (7.20) yields an exponential potential

$$V(\phi) = V_0 e^{-\kappa\lambda\phi} . \quad (7.23)$$

From Eq. (7.22) this potential corresponds to $\Gamma = 1$. In this case the autonomous equations (7.18) and (7.19) are closed. The cosmological dynamics can be well understood by studying fixed points of the system [37], as we will see below.

If Γ is constant but λ is not, we have to solve Eq. (7.21) as well as Eqs. (7.18) and (7.19). For the power-law potential, $V(\phi) = M^{4+n}\phi^{-n}$ ($n > 0, \phi > 0$), we have that $\Gamma = (n + 1)/n > 1$ and $x_1 > 0$, in which case the quantity λ (> 0) decreases from Eq. (7.21). Of course, for general field potentials, Γ is not necessarily constant. In such cases we need to obtain the field ϕ as a function of N by solving Eqs. (7.15) and (7.19) together with the use of the relation $\kappa\sqrt{V} = \sqrt{3}Hx_2$. Then the evolution of the variable $\lambda = \lambda(\phi)$ is known accordingly.

In the following we first discuss cosmological dynamics for the exponential potential given in Eq. (7.23) and then proceed to the case of non-constant λ .

7.2.1 Exponential potential

We can derive fixed points of the system by setting $dx_1/dN = dx_2/dN = 0$ in Eqs. (7.18) and (7.19). The fixed points are in general the solution of the dynamical system and give a first qualitative description of the phase space. As we discuss below they can be classified according to their stability properties. If there are no

singularities or strange attractors, the trajectories with respect to $x_1(N)$ and $x_2(N)$, in general to be obtained numerically, run from unstable fixed points to stable points, coasting along “saddle” points.

When λ is constant they are given by

- (a) $(x_1, x_2) = (0, 0)$, $\Omega_\phi = 0$, $w_{\text{eff}} = w_M$, w_ϕ : undefined.
- (b1) $(x_1, x_2) = (+1, 0)$, $\Omega_\phi = 1$, $w_{\text{eff}} = 1$, $w_\phi = 1$.
- (b2) $(x_1, x_2) = (-1, 0)$, $\Omega_\phi = 1$, $w_{\text{eff}} = 1$, $w_\phi = 1$.
- (c) $(x_1, x_2) = (\lambda/\sqrt{6}, [1 - \lambda^2/6]^{1/2})$, $\Omega_\phi = 1$, $w_{\text{eff}} = -1 + \lambda^2/3$, $w_\phi = -1 + \lambda^2/3$.
- (d) $(x_1, x_2) = (\sqrt{3/2}(1 + w_M)/\lambda, [3(1 - w_M^2)/2\lambda^2]^{1/2})$, $\Omega_\phi = 3(1 + w_M)/\lambda^2$, $w_{\text{eff}} = w_M$, $w_\phi = w_M$.

The point (a) is a fluid-dominated solution ($\Omega_M = 1$). The kinetic energy of quintessence is dominant for the points (b1) and (b2), in which case ρ_ϕ decreases rapidly ($\rho_\phi \propto a^{-6}$) relative to the background density. The point (c) corresponds to a scalar-field-dominated solution, which exists for $\lambda^2 < 6$. The cosmic acceleration is realized if $w_{\text{eff}} < -1/3$, i.e. $\lambda^2 < 2$. In the limit that $\lambda \rightarrow 0$ (i.e. $V(\phi) \rightarrow V_0$) we recover the equation of state of cosmological constant ($w_{\text{eff}} = w_\phi = -1$). The point (d) is the so-called scaling solution [37] in which the ratio Ω_ϕ/Ω_M is a non-zero constant. The existence of the scaling solution demands the condition $\lambda^2 > 3(1 + w_M)$ from the requirement $\Omega_\phi < 1$. Since $w_\phi = w_M$ for scaling solutions, it is not possible to realize cosmic acceleration unless the matter fluid has an unusual equation of state ($w_M < -1/3$).

In order to find the stability about the fixed points $(x_1^{(c)}, x_2^{(c)})$ derived above, we consider linear perturbations $(\delta x_1, \delta x_2)$ as follows:

$$x_1 = x_1^{(c)} + \delta x_1, \quad x_2 = x_2^{(c)} + \delta x_2. \quad (7.24)$$

Linearizing Eqs. (7.18) and (7.19) leads to the first-order differential equations

$$\frac{d}{dN} \begin{pmatrix} \delta x_1 \\ \delta x_2 \end{pmatrix} = \mathcal{M} \begin{pmatrix} \delta x_1 \\ \delta x_2 \end{pmatrix}, \quad (7.25)$$

where $\mathcal{M}$ is a 2×2 matrix whose components depend upon $x_1^{(c)}$ and $x_2^{(c)}$. The eigenvalues of the matrix $\mathcal{M}$ are given by

$$\mu_{1,2} = \frac{1}{2} \left[a_{11} + a_{22} \pm \sqrt{(a_{11} + a_{22})^2 - 4(a_{11}a_{22} - a_{12}a_{21})} \right], \quad (7.26)$$

where a_{ij} are the components of the matrix. The general linearized solution around each fixed point can be written then as

$$x_i = x_i^{(c)} + \alpha_{i1} e^{\mu_{11}N} + \alpha_{i2} e^{\mu_{22}N}, \quad (7.27)$$

where α_{i1} and α_{i2} are coefficients obtained from the eigenvectors. The eigenvalues determine therefore the behavior of solutions near the fixed points.

Denoting the determinant of the matrix $\mathcal{M}$ as $\mathcal{D} \equiv (a_{11} + a_{22})^2 - 4(a_{11}a_{22} - a_{12}a_{21})$, the stability of the fixed points can be generally classified in the following way:

- (i) Stable node: $\mathcal{D} > 0$ and $\mu_1 < 0, \mu_2 < 0$.
- (ii) Unstable node: $\mathcal{D} > 0$ and $\mu_1 > 0, \mu_2 > 0$.
- (iii) Saddle point: $\mathcal{D} > 0$ and $\mu_1 < 0, \mu_2 > 0$ (or $\mu_1 > 0$ and $\mu_2 < 0$).
- (iv) Stable spiral: $\mathcal{D} < 0$ and the real parts of μ_1 and μ_2 are negative.
- (v) Unstable spiral: $\mathcal{D} < 0$ and the real parts of μ_1 and μ_2 are positive.

If $\mathcal{D} = 0$ the matrix $\mathcal{M}$ is singular and the system becomes effectively one-dimensional around the fixed point. This classification can be extended to more dimensions: a fixed point is stable if all the real parts of the eigenvalues are negative, unstable if they are all positive, and a saddle when there are negative and positive real parts. If an eigenvalue vanishes then the stability can be established expanding to higher orders. We will use this dynamical system approach repeatedly in the course of this book.

The eigenvalues of the above fixed points are given by (see problem 7.1)

- Point (a): $\mu_1 = -\frac{3}{2}(1 - w_M), \quad \mu_2 = \frac{3}{2}(1 + w_M)$.
- Point (b1): $\mu_1 = 3 - \frac{\sqrt{6}}{2}\lambda, \quad \mu_2 = 3(1 - w_M)$.
- Point (b2): $\mu_1 = 3 + \frac{\sqrt{6}}{2}\lambda, \quad \mu_2 = 3(1 - w_M)$.
- Point (c): $\mu_1 = \frac{1}{2}(\lambda^2 - 6), \quad \mu_2 = \lambda^2 - 3(1 + w_M)$.
- Point (d): $\mu_{1,2} = -\frac{3(1-w_M)}{4} \left[1 \pm \sqrt{1 - \frac{8(1+w_M)[\lambda^2 - 3(1+w_M)]}{\lambda^2(1-w_M)}} \right]$.

Let us consider a realistic case in which the equation of state of the fluid is in the region $0 \leq w_M < 1$. Then the stability of the fixed points is summarized as follows.

- Point (a): Saddle.
- Point (b1): Unstable node for $\lambda < \sqrt{6}$ and saddle point for $\lambda > \sqrt{6}$.
- Point (b2): Unstable node for $\lambda > -\sqrt{6}$ and saddle point for $\lambda < -\sqrt{6}$.
- Point (c): Stable node for $\lambda^2 < 3(1 + w_M)$ and saddle point for $3(1 + w_M) < \lambda^2 < 6$.
- Point (d): Saddle for $\lambda^2 < 3(1 + w_M)$, stable node for $3(1 + w_M) < \lambda^2 < \frac{24(1+w_M)^2}{7+9w_M}$ and stable spiral for $\lambda^2 > \frac{24(1+w_M)^2}{7+9w_M}$.

The radiation ($w_M = 1/3$) and matter ($w_M = 0$) dominated epochs can be realized either by the point (a) or (d). When $\lambda^2 > 3(1 + w_M)$ the solutions approach the stable scaling fixed point (d) instead of the point (a). In this case, however, the solutions do not exit from the scaling era ($\Omega_\phi = \text{constant}$) to connect to the accelerated epoch. In order to give rise to tracking behavior in which Ω_ϕ evolves to catch

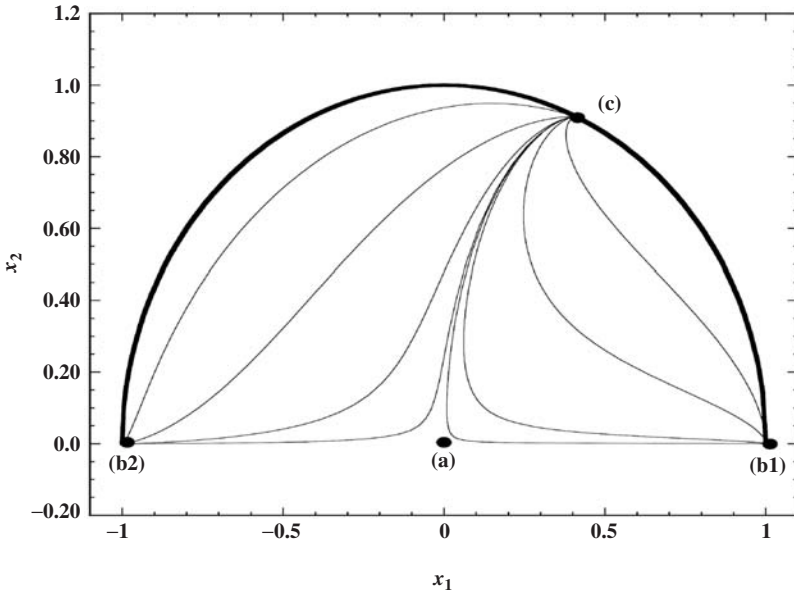


Figure 7.1 The trajectories of solutions for the exponential potential (7.23) with model parameters $\lambda = 1$ and $w_M = 0$. In this case the attractor is the accelerated point (c) $(x_1, x_2) = (0.4082, 0.9129)$. The matter point (a) is a saddle, whereas the points (b1) and (b2) are unstable nodes. The thick curve is the border of the allowed region characterized by $x_2 = \sqrt{1 - x_1^2}$.

up with Ω_M , we require that the slope of the potential gradually decreases. This can be realized by the field potential in which λ gets smaller with time (such as $V(\phi) = M^{4+n}\phi^{-n}$). We will discuss this case in Section 7.2.2. It is worth mentioning that the exponential potential corresponds to the border that separates regions where such tracking behavior occurs from those where it does not.

The point (c) is the only fixed point giving rise to a stable accelerated attractor for $\lambda^2 < 2$. When $\lambda^2 < 2$, a physically meaningful solution (d) does not exist because $\Omega_\phi > 1$ for both radiation and matter fluids. In this case the radiation- and matter-dominated epochs are realized by the point (a). Note that when λ is close to 0 the solution starting from the point (a) and approaching the point (c) is not much different from the cosmological constant scenario. Nevertheless, since the equation of state of the attractor is given by $w_\phi = -1 + \lambda^2/3$, we can still find a difference from $w_\phi = -1$.

In Fig. 7.1 we plot the trajectories of solutions in the (x_1, x_2) plane for $\lambda = 1$ and $w_M = 0$. Since $\Omega_M \geq 0$ in Eq. (7.13), the allowed region corresponds to $0 \leq x_2 \leq \sqrt{1 - x_1^2}$. The kinetic-energy-dominated points (b1) and (b2) are unstable in this case. Since the matter point (a) is a saddle, the solutions starting from

$x_2 \ll 1$ temporarily approach this fixed point. The trajectories finally approach the accelerated fixed point (c), because this is stable for $\lambda^2 < 3$.

7.2.2 Other potentials

If λ is not constant, we need to solve Eq. (7.21) to know the evolution of λ . In this case, the fixed points derived in the constant λ case can be regarded as “instantaneous” fixed points changing in time [186, 187], provided that the time scale for the variation of λ is much less than H^{-1} .

Let us consider the “freezing” models of quintessence without a potential minimum (such as $V(\phi) = M^{4+n}\phi^{-n}$). We then have $\lambda > 0$ and $x_1 > 0$ for $V_{,\phi} < 0$ and $\lambda < 0$ and $x_1 < 0$ for $V_{,\phi} > 0$. If the condition

$$\Gamma = \frac{V_{,\phi\phi}}{V_{,\phi}^2} > 1 \quad (7.28)$$

is satisfied, the absolute value of λ decreases toward 0 irrespective of the sign of $V_{,\phi}$. This means that the solutions finally approach the accelerated “instantaneous” point (c) even if λ^2 is larger than 2 during radiation and matter eras. The condition (7.28) is the so-called tracking condition under which the field density eventually catches up that of the background fluid.

The condition (7.28) can also be derived in the following way [39]. We first define the following quantity

$$x \equiv \frac{1 + w_\phi}{1 - w_\phi} = \frac{\dot{\phi}^2}{2V}. \quad (7.29)$$

Taking the derivative of x in terms of N and using the definition (7.14), we find

$$\frac{V_{,\phi}}{\kappa V} = \pm \sqrt{\frac{3(1 + w_\phi)}{\Omega_\phi}} \left(1 + \frac{1}{6} \frac{d \ln x}{dN} \right), \quad (7.30)$$

where the plus and minus signs correspond to the cases $\dot{\phi} < 0$ and $\dot{\phi} > 0$, respectively. Differentiating Eq. (7.30) with respect to ϕ , we get the following relation (see problem 7.2):

$$\begin{aligned} \Gamma = 1 + & \frac{3(1 - \Omega_\phi)(w_M - w_\phi)}{(1 + w_\phi)(6 + y')} - \frac{y'}{(1 + w_\phi)(6 + y')(1 + x)} \\ & - \frac{2y''}{(1 + w_\phi)(6 + y')^2}, \end{aligned} \quad (7.31)$$

where $y' \equiv d \ln x / dN$. Let us consider the evolution during the radiation- and matter-dominated epochs where Ω_ϕ can be negligible relative to 1. If Γ varies

slowly in time, Eq. (7.31) implies that there is a solution in which w_ϕ is nearly constant and its derivatives (y' and y'') are negligible. Hence the equation of state of quintessence is nearly constant:

$$w_\phi \simeq \frac{w_M - 2(\Gamma - 1)}{1 + 2(\Gamma - 1)}. \quad (7.32)$$

The exponential potential corresponds to $\Gamma = 1$, giving the scaling solution with $w_\phi = w_M$. If $\Gamma > 1$, we have $w_\phi < w_M$ so that the quintessence energy density evolves more slowly than the background energy density. Hence the tracking solution can be realized under the condition (7.28) for Γ nearly constant ($|d(\Gamma - 1)/dN| \ll |\Gamma - 1|$).

Let us consider the inverse power-law potential $V(\phi) = M^{4+n}\phi^{-n}$ ($n > 0$). Since $\Gamma = (n + 1)/n$ in this case, the tracking condition (7.28) is automatically satisfied. The epoch of the late-time cosmic acceleration is quantified by the condition $\lambda^2 = V_\phi^2/(\kappa^2 V^2) < 2$, i.e.

$$\phi > \frac{n}{4\sqrt{\pi}} m_{\text{pl}}, \quad (7.33)$$

which is independent of the mass scale M . The field value at the onset of the accelerated expansion is of the order of the Planck mass for $n = \mathcal{O}(1)$. From the Friedmann equation (7.7) one can estimate the present potential energy of quintessence to be $V(\phi_0) \approx H_0^2 m_{\text{pl}}^2$, where $\phi_0 \approx m_{\text{pl}}$. Then the mass M is constrained to be

$$M \approx \left(\frac{H_0}{m_{\text{pl}}} \right)^{\frac{2}{4+n}} m_{\text{pl}} \approx 10^{-\frac{46-19n}{4+n}} \text{ GeV}, \quad (7.34)$$

where we have used $H_0 \approx 10^{-42} \text{ GeV}$. For $n = 2$ and $n = 4$ we have that $M \approx 10^{-1} \text{ GeV}$ and $M \approx 10^4 \text{ GeV}$, respectively. These energy scales can be compatible with those appearing in particle physics.

Let us consider the thawing models of quintessence in which the field was frozen in the past and started to move only recently. For example, in the case of the potential $V(\phi) = M^{4-n}\phi^n$ ($n > 0$), we have $\Gamma = (n - 1)/n < 1$ and hence the model does not satisfy the tracking condition (7.28). Since $|\lambda| = (n/\sqrt{8\pi})(m_{\text{pl}}/|\phi|)$, the late-time acceleration occurs only in the region $|\phi| > nm_{\text{pl}}/4\sqrt{\pi}$. This shows that the initial field displacement ϕ_i and the field mass m_ϕ about the potential minimum are crucially important to get the cosmic acceleration. As long as $|\phi_i| \gtrsim m_{\text{pl}}$ and $m_\phi \lesssim H_0$, the Universe enters a temporary phase of accelerated expansion. The cosmic acceleration ends after the field $|\phi|$ drops down to the order of m_{pl} . The potential $V(\phi) = M^4 \cos^2(\phi/f)$ also exhibits similar cosmic expansion history. The situation is different for the model in which the potential has a non-vanishing energy V_0 at the potential minimum, e.g., $V(\phi) = V_0 + M^{4-n}\phi^n$ with $V_0 > 0$. In

this case $|\lambda|$ eventually approaches 0 so that the potential energy at $\phi = 0$ can be responsible for dark energy.

7.2.3 Tracker solutions

Beside fixed points, phase spaces can be characterized also by special trajectories that “attract” other trajectories. Tracker solutions have approximately constant w_ϕ and Ω_ϕ along these special attractors. A wide range of initial conditions converge to a common, cosmic evolutionary tracker.

In this section we shall discuss tracker solutions in detail. To be concrete we consider the inverse power-law potential

$$V(\phi) = M^{4+n} \phi^{-n}, \quad (n > 0). \quad (7.35)$$

We study the evolution of the scalar field in the region $\phi > 0$, i.e. $V_{,\phi} < 0$ and $\dot{\phi} > 0$. We take into account both radiation (energy density ρ_r) and non-relativistic matter (energy density ρ_m) together with the quintessence field. In this case the total energy density ρ_M and the pressure P_M of fluids in Eqs. (7.7) and (7.8) are given by $\rho_M = \rho_r + \rho_m$ and $P_M = \rho_r/3$, respectively. In addition to the variables x_1 and x_2 defined in Eq. (7.12) we introduce another variable: $x_3 \equiv \kappa \sqrt{\rho_r}/(\sqrt{3}H)$. Then the density parameters for quintessence, radiation, and non-relativistic matter are

$$\Omega_\phi = x_1^2 + x_2^2, \quad \Omega_r = x_3^2, \quad \Omega_m = 1 - x_1^2 - x_2^2 - x_3^2. \quad (7.36)$$

The effective equation of state defined in Eq. (4.96) reads

$$w_{\text{eff}} = x_1^2 - x_2^2 + x_3^2/3. \quad (7.37)$$

The equation of state of quintessence is the same as Eq. (7.6).

The equations for x_1 , x_2 , and x_3 are

$$\frac{dx_1}{dN} = -3x_1 + \frac{\sqrt{6}}{2} \lambda x_2^2 + \frac{1}{2} x_1 (3 + 3x_1^2 - 3x_2^2 + x_3^2), \quad (7.38)$$

$$\frac{dx_2}{dN} = -\frac{\sqrt{6}}{2} \lambda x_1 x_2 + \frac{1}{2} x_2 (3 + 3x_1^2 - 3x_2^2 + x_3^2), \quad (7.39)$$

$$\frac{dx_3}{dN} = -2x_3 + \frac{1}{2} x_3 (3 + 3x_1^2 - 3x_2^2 + x_3^2). \quad (7.40)$$

Using the fact that $\Gamma = (n + 1)/n$ for the potential (7.35), the equation for λ reads

$$\frac{d\lambda}{dN} = -\sqrt{6} \frac{\lambda^2}{n} x_1. \quad (7.41)$$

Note that $\lambda > 0$ because we are considering the case $V_{,\phi} < 0$. Since $x_1 > 0$ (because $\dot{\phi} > 0$), the r.h.s. of Eq. (7.41) is negative so that λ decreases with time. The equations (7.38)–(7.41) are the autonomous equations to be solved numerically.

From Eq. (7.32) the equation of state of quintessence in the tracking regime is given by

$$w_\phi \simeq \frac{nw_M - 2}{n + 2}. \quad (7.42)$$

If the tracking occurs during the matter-dominated epoch ($w_M = 0$), then $w_\phi \simeq -2/(n + 2)$. From Eq. (7.30) the following relation holds

$$\frac{1}{6} \frac{d \ln x}{dN} = \Delta(t) - 1, \quad \text{where} \quad \Delta(t) \equiv \lambda \sqrt{\frac{\Omega_\phi}{3(1 + w_\phi)}}. \quad (7.43)$$

From the definition of x in Eq. (7.29) we also obtain

$$\frac{1}{6} \frac{d \ln x}{dN} = \frac{1}{3(1 - w_\phi^2)} \frac{dw_\phi}{dN}. \quad (7.44)$$

Since w_ϕ is nearly constant for tracker solutions, it follows from Eqs. (7.43) and (7.44) that $\Delta \simeq 1$. Hence the tracker solution is characterized by

$$\Omega_\phi \simeq \frac{3(1 + w_\phi)}{\lambda^2}, \quad (7.45)$$

where w_ϕ is given in Eq. (7.42). Recall that the scaling fixed point (d) for constant λ corresponds to $\Omega_\phi = 3(1 + w_M)/\lambda^2$ and $w_\phi = w_M$. In this case the tracker solution (7.45) recovers the scaling solution in the regime $\lambda^2 > 3(1 + w_M)$ (under which the scaling solution is stable). The accelerated fixed point (c) for constant λ corresponds to $\Omega_\phi = 1$ and $w_\phi = -1 + \lambda^2/3$. The tracker solution (7.45) also covers this case and the accelerated solution is stable for $\lambda^2 < 3(1 + w_M)$. Hence the tracker solution can be regarded as a stable attractor. For constant λ the stable scaling solution (d) does not exit to the accelerated attractor (c), but for decreasing λ the transition to the stable accelerated phase occurs through the tracking solution.

Unlike the cosmological constant scenario we are interested in the case where the energy density of quintessence is not completely negligible relative to the background energy density even during the radiation era. Given the fact that λ grows toward the past, the parameter Δ defined in Eq. (7.43) can be much larger than unity at the beginning of the radiation era. Under the initial conditions $\Delta(t_i) \gg 1$ the Universe finally converges to the tracking solution in the following way [38].

- (i) Since $\Delta \gg 1$ at the initial stage, it follows from Eqs. (7.43) and (7.44) that w_ϕ is driven to the maximum value, $w_\phi \rightarrow 1$. Since this corresponds to the stage in which the kinetic energy of quintessence is dominant, the field rolls down the potential quickly.

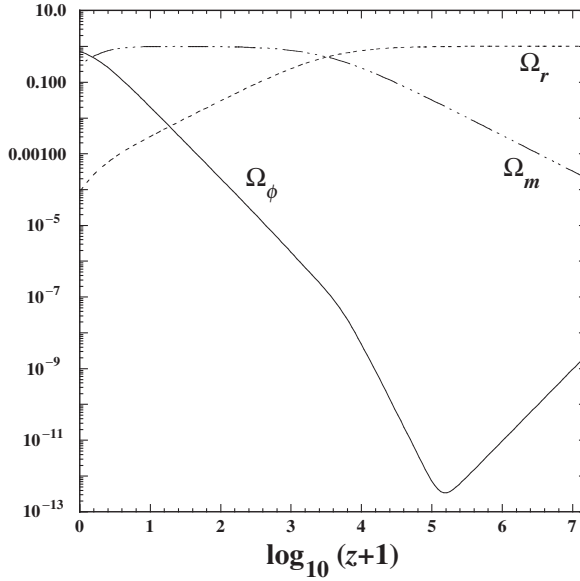


Figure 7.2 Evolution of Ω_ϕ , Ω_m , and Ω_r for the model $V(\phi) = M^5\phi^{-1}$ versus the redshift z . Initially Ω_ϕ rapidly decreases because the field equation of state is given by $w_\phi \simeq 1$. This is followed by the phase with a nearly frozen scalar field so that Ω_ϕ begins to grow. Finally the solution enters the tracking regime in which the field energy density tracks the background fluid density. Initial conditions are chosen to be $x_1 = 5.0 \times 10^{-5}$, $x_2 = 1.0 \times 10^{-8}$, $x_3 = 0.9999$, and $\lambda = 10^9$ at $\log_{10}(z+1) = 7.21$.

- (ii) During the rapidly rolling phase of the scalar field both λ and Ω_ϕ decrease, which leads to the decrease of Δ . The quantity Δ eventually reaches a tracker value, $\Delta = 1$. However, Eqs. (7.43) and (7.44) imply that $dw_\phi/dN > 0$ for $\Delta > 1$. Hence w_ϕ stays around $w_\phi = 1$ up to the moment at which Δ crosses 1. In this situation the kinetic energy of quintessence is too large for ϕ to join the tracker solution. The field overshoots the tracker solution with a fast rolling down along the potential.
- (iii) After the overshooting of the tracker solution, one has $\Delta < 1$ so that w_ϕ begins to decrease and approaches the value -1 . Once w_ϕ becomes close to -1 , the field is almost frozen so that λ becomes close to 0. However Ω_ϕ starts to grow and hence this leads to the increase of Δ . As long as $\Delta < 1$, w_ϕ remains around -1 because the condition $dw_\phi/dN < 0$ is satisfied.
- (iv) Once Δ becomes larger than 1, w_ϕ begins to increase from -1 . Then the field starts to evolve again along the potential hill. The sign of Eq. (7.43) changes again. Consequently Δ approaches unity after a few oscillations and the field enters the tracking regime.

In Fig. 7.2 we plot an example for the evolution of Ω_ϕ , Ω_m , and Ω_r for the potential $V(\phi) = M^5\phi^{-1}$. The present epoch ($z = 0$ and $N = N_0$) is identified as

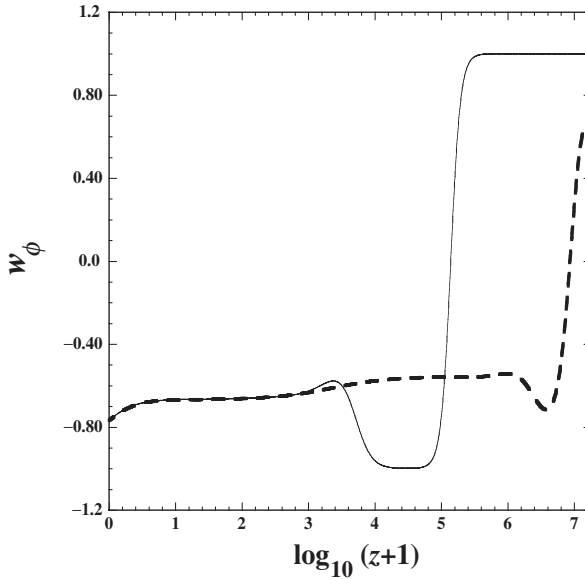


Figure 7.3 Evolution of w_ϕ for the model $V(\phi) = M^5 \phi^{-1}$ with two different initial conditions. The solid curve corresponds to the case shown in Fig. 7.2. The dashed curve corresponds to the case with initial conditions $x_1 = 1.0 \times 10^{-8}$, $x_2 = 1.0 \times 10^{-8}$, $x_3 = 0.9999$, and $\lambda = 10^9$ at $\log_{10}(z+1) = 7.23$. Both curves finally converge to the tracking solution whose equation of state is given by $w_\phi \simeq -2/3$ during the matter-dominated epoch ($10 \lesssim z \lesssim 10^3$).

$\Omega_\phi = 0.72$. The relation between z and N is given by $z = e^{N_0 - N} - 1$. Here the Universe is initially in the regime $\Delta \gg 1$ and $w_\phi \simeq 1$ so that the quintessence energy density rapidly decreases (see also the solid curve in Fig. 7.3). The equation of state w_ϕ starts to decrease for $\Delta < 1$ and stays around $w_\phi \simeq -1$ for the redshift $10^4 \lesssim z \lesssim 10^5$. The field is nearly frozen in this regime so that Ω_ϕ grows rapidly. The equation of state w_ϕ begins to grow from -1 and then it approaches a nearly constant value after small oscillations. Note that the tracking solution in this case corresponds to the redshift $z \lesssim 10^3$, i.e. the matter-dominated epoch ($w_M = 0$). Hence from Eq. (7.42) it follows that the equation of state in the tracker regime is given by $w_\phi \simeq -2/3$. In fact the numerical simulation in Fig. 7.3 shows that w_ϕ is close to $-2/3$ for the redshift $10 \lesssim z \lesssim 10^3$. Recall that in order to derive Eq. (7.32) from Eq. (7.31) we have used the approximation that Γ is nearly constant. Since Γ begins to vary once the contribution of the quintessence energy density becomes important, the equation of state w_ϕ starts to decrease in the low-redshift region ($z \lesssim 10$).

In Fig. 7.3 we also show the evolution of w_ϕ for another initial condition as a dashed curve. Since the kinetic energy of quintessence does not dominate over its potential energy at the initial stage, the early evolution of w_ϕ is different from the

case of the solid curve discussed above. However this solution also enters the tracking regime so that the late-time evolution of w_ϕ is the same as the “overshooting” case. Hence the solutions finally approach the tracking attractor independent of initial conditions. This property is attractive to alleviate the coincidence problem. However since Ω_ϕ grows in the future to unity it remains to be explained why it equals Ω_m near the present epoch.

7.3 Early dark energy

One of the differences between quintessence and the Λ CDM model is that the energy density of the former can contribute to the total energy density even in the early epoch of the cosmological evolution. It is then possible to place constraints on quintessence models by studying the evolution of the field during radiation and matter dominated epochs. For the exponential potential (7.23) we have shown that there exists the scaling solution (d) with $\Omega_\phi = 3(1 + w_M)/\lambda^2$ and $w_\phi = w_M$, which is stable for $\lambda^2 > 3(1 + w_M)$. In this case the solutions approach the scaling fixed point (d) rather than the saddle fluid dominated point (a) in radiation- and matter-dominated epochs. For scaling solutions the energy density of the field decreases in proportion to the background fluid. Since $\Omega_\phi = 4/\lambda^2$ and $\Omega_\phi = 3/\lambda^2$ during radiation and matter eras, respectively, the existence of stable scaling solutions requires the condition $\lambda^2 > 4$.

Of course the solutions need to finally exit from the scaling regime to the epoch dominated by dark energy. Since the exponential potential $V(\phi) = V_0 e^{-\kappa\lambda\phi}$ with $\lambda^2 > 4$ is too steep to give rise to the late-time acceleration, we require that the potential becomes shallow or it has a minimum to slow down the movement of the scalar field. One explicit model is the double exponential potential [188]

$$V(\phi) = V_1 e^{-\kappa\lambda\phi} + V_2 e^{-\kappa\mu\phi} . \quad (7.46)$$

Let us consider the case $\lambda^2 > 4$ and $\mu^2 < 2$. During radiation and matter eras the field is in a scaling regime driven by the steep potential $V(\phi) \simeq V_1 e^{-\kappa\lambda\phi}$. The solutions exit from the scaling era to the epoch of cosmic acceleration once the exponential potential with the slope μ becomes important. An attractive point of this model is that the solutions are temporarily trapped by the scaling solution irrespective of the initial conditions of the field, see Fig. 7.4. Other potentials, as e.g., $V(\phi) = V_0 [\cosh(\lambda\phi) - 1]^p$, exhibit similar properties [189].

This scenario can be generalized to an N -scalar quintessence whose potential is the sum of exponential potentials with *multiple* fields ($\phi_1, \phi_2, \dots, \phi_N$):

$$V = \sum_{i=1}^N V_i e^{-\kappa\lambda_i\phi_i} , \quad (7.47)$$

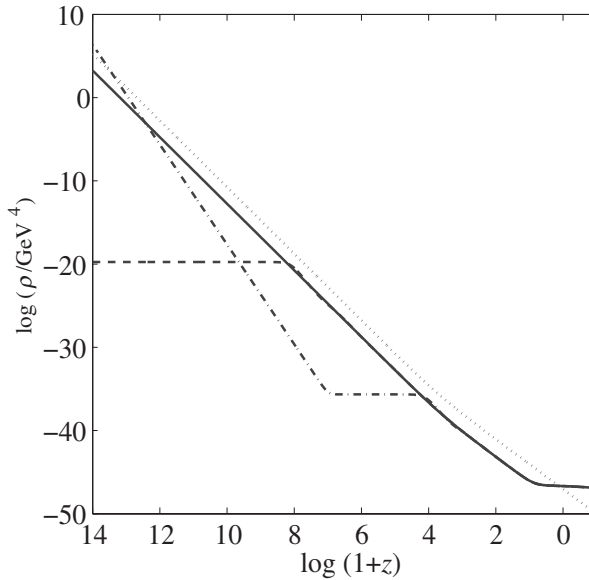


Figure 7.4 Evolution of the quintessence energy density ρ_ϕ and the background fluid density $\rho_r + \rho_m$ for the model (7.46) with parameters $\lambda = 20$ and $\mu = 0.5$. Several initial conditions are chosen to admit an $\Omega_\phi = 0.7$ flat Universe today. The solid line shows the evolution which emerges from equipartition at the end of inflation and the dotted line represents the evolution of $\rho_r + \rho_m$. From Ref. [188].

where V_i and λ_i are constants. This model leads to the so-called assisted inflation phenomenon where several scalar fields cooperate to drive the cosmic acceleration even if none is able to individually. In the context of inflation Liddle *et al.* [190] showed that the N -scalar fields evolve to give dynamics matching a single-field model with an effective slope λ_{eff} defined by (see problem 7.3)

$$\frac{1}{\lambda_{\text{eff}}^2} = \sum_{i=1}^N \frac{1}{\lambda_i^2}. \quad (7.48)$$

The solutions approach the stable accelerated attractor for $\lambda_{\text{eff}}^2 < 2$. If there is a large number of exponential potentials with different slopes and initial conditions, more and more fields would join the assisted quintessence attractor as the Universe expands so that λ_{eff} reduces. Even if one of the exponential potentials with a large slope dominates at the early epoch of the cosmological evolution, the solutions eventually exit from the (nearly) scaling regime to the accelerated attractor [191, 192, 193].

For the quintessence potentials given by Eqs. (7.46) and (7.47), the field energy density is non-negligible relative to the background density even in the early epoch of cosmological evolution. There is a class of models dubbed quintessential

inflation [194] in which a single scalar field ϕ is responsible for both inflation and dark energy. If the field has a potential minimum just after inflation, it decays to radiation during the reheating stage. For the potential, $V(\phi) = \Lambda + m_\phi^2 \phi^2/2$, one can realize inflation for $m_\phi \simeq 10^{-6} m_{\text{pl}}$ [94] and the late-time cosmic acceleration for $\Lambda \simeq 10^{-123} m_{\text{pl}}^4$. However, the field is stabilized at the potential minimum soon after inflation. Hence this model cannot be distinguished from the cosmological constant scenario.

A potential without a minimum can be described as [194]

$$V(\phi) = \begin{cases} \lambda(\phi^4 + M^4) & (\phi < 0), \\ \lambda M^4 / [1 + (\phi/M)^n] & (\phi \geq 0). \end{cases} \quad (7.49)$$

In the regime of a negative scalar field with $|\phi| \gg M$, this corresponds to the self-coupling chaotic inflation potential: $V(\phi) \simeq \lambda \phi^4$. The COBE normalization constrains the coupling λ to be $\lambda \approx 10^{-13}$ [94]. In the positive field regime with $\phi \gg M$, Eq. (7.49) reduces to the inverse power-law potential: $V(\phi) \simeq \tilde{M}^{4+n} \phi^{-n}$ with $\tilde{M} \equiv \lambda^{1/(4+n)} M$. From the requirement (7.34) of the cosmic acceleration today, the mass scale $\tilde{M}$ is constrained to be $\tilde{M} \approx 10^{-\frac{46-19n}{4+n}} \text{ GeV}$. Since the potential does not possess a minimum, the reheating after inflation should proceed through gravitational particle production rather than through direct matter couplings between the field ϕ and other scalar fields. This mechanism is generally inefficient to lead to a complete decay of the inflaton field to radiation [195]. As we will see below, it is problematic if the quintessence energy density is comparable to the radiation density at the epoch of nucleosynthesis. We note, however, that the instant preheating scenario proposed by Felder *et al.* [196] may alleviate the reheating problem of quintessential inflation through the rapid decay of ϕ into another field χ with an interaction $(1/2)g^2\phi^2\chi^2$.

Let us discuss observational constraints on the density parameter Ω_ϕ in the early epoch of cosmological evolution. The tightest bound comes from the Big Bang Nucleosynthesis (BBN), which provides an upper bound on Ω_ϕ from a number of observations around the temperature $T = 1 \text{ MeV}$. The presence of the scaling scalar field leads to a larger expansion rate of the Universe. This leads to the change of the ratio of neutrons to protons at freeze-out and hence the abundances of light elements such as helium (He) and deuterium (D) are modified.

Using the observed values for the mass fraction of helium, $Y_{\text{He}} = 0.244 \pm 0.002$, and a relative abundance for deuterium, $D/H = (3.0 \pm 0.4) \times 10^{-5}$, one can find constraints on the parameter space $(\Omega_\phi^{\text{BBN}} h^2, \Omega_b^{(0)} h^2)$. Here Ω_ϕ^{BBN} is the density parameter of quintessence at the BBN epoch. As we see in Fig. 7.5, the bound on Ω_ϕ^{BBN} is given by [197]

$$\Omega_\phi^{\text{BBN}} < 0.045, \quad (7.50)$$

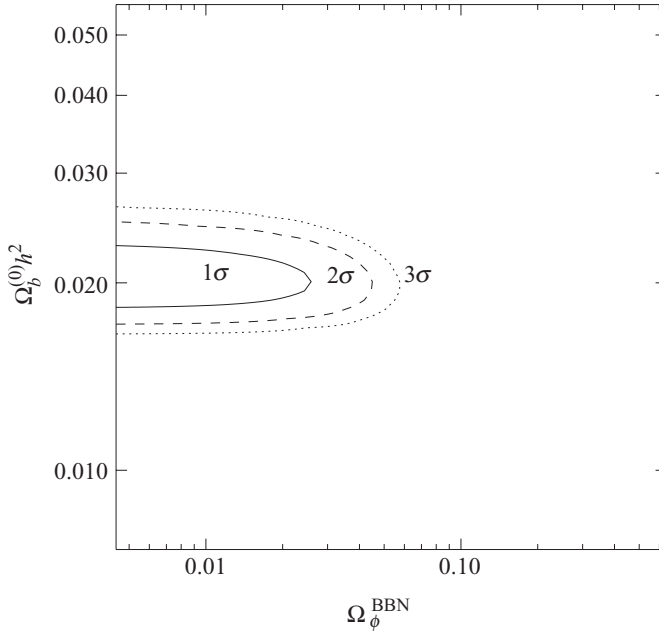


Figure 7.5 Observational contours (1, 2, 3 σ confidence levels) on the parameters $(\Omega_\phi^{\text{BBN}}, \Omega_b^{(0)}h^2)$ derived from the abundances $Y_{\text{He}} = 0.244 \pm 0.002$ and $D/H = (3.0 \pm 0.4) \times 10^{-5}$. From Ref. [197].

at the 95% confidence level. Note that the bound on the density parameter Ω_ϕ^{BBN} can be interpreted in terms of the maximum number of relativistic degrees of freedom ΔN_{eff} coming from the scalar field [33]:

$$\Omega_\phi^{\text{BBN}} < \frac{7\Delta N_{\text{eff}}/4}{10.75 + 7\Delta N_{\text{eff}}/4}, \quad (7.51)$$

where 10.75 in the denominator is the number of effective degrees of freedom in the standard model. Then the constraint (7.50) corresponds to the bound $\Delta N_{\text{eff}} < 0.3$. If we use a more conservative bound $0.9 < \Delta N_{\text{eff}} < 1.5$ adopted in Refs. [198, 199], the constraint on Ω_ϕ^{BBN} becomes weaker: $\Omega_\phi^{\text{BBN}} < 0.13\text{--}0.2$ [33].

The presence of scaling radiation and matter eras also modifies the CMB power spectrum. There are several effects on CMB [17, 197]: (i) the change of the radiation–matter equality modifies the structure of CMB acoustic peaks, (ii) the change of the matter content of the Universe leads to a shift for the position of peaks, (iii) the modified evolution of gravitational potential changes the large-scale CMB spectrum through the ISW effect. Using the Boomerang and DASI data, Bean *et al.* [197] derived the constraint $\Omega_\phi^{\text{BBN}} < 0.39$ at the 95% confidence level, which is weaker than the bound (7.50).

Taking the bound (7.50), the constraint on λ for the model (7.46) is given by

$$\lambda > 9.4. \quad (7.52)$$

If we take a conservative bound $\Omega_\phi^{\text{BBN}} < 0.2$, the constraint becomes $\lambda > 4.5$.

In Section 8.3 we will see that another way to realize an early dark energy scenario is to couple dark energy and dark matter.

7.4 Quintessence potentials in particle physics

We now turn to the construction of quintessence models under the framework of particle physics. It is generally not so easy to construct viable quintessence models because the energy scale of dark energy is too low ($\rho_\phi \simeq 10^{-123} m_{\text{pl}}^4$) relative to typical scales appearing in particle physics. In order to lead to the cosmic acceleration today we require that the potential is flat enough to satisfy the condition $|\eta_s| \lesssim 1$, where η_s is the slow-roll parameter defined in Eq. (7.10). This corresponds to the condition

$$\left| \frac{M_{\text{pl}}^2 V_{,\phi\phi}}{V} \right| \lesssim 1, \quad (7.53)$$

where M_{pl} is the reduced Planck mass. Hence the quintessence mass squared, $m_\phi^2 \equiv V_{,\phi\phi}$, needs to satisfy

$$|m_\phi^2| \lesssim V_0/M_{\text{pl}}^2 \simeq H_0^2, \quad \rightarrow \quad |m_\phi| \lesssim H_0 \approx 10^{-33} \text{ eV}, \quad (7.54)$$

where V_0 is the potential energy today. This means that to be compatible with the present cosmic acceleration the field mass must be extremely small. Such a light scalar field can give rise to couplings to ordinary matter which should lead to observable long-range forces [36]. Moreover the mass is unstable against radiative corrections in the absence of supersymmetry. The radiative corrections can disrupt the flatness of the quintessence potential required for the cosmic acceleration [200].

Despite these difficulties it is not entirely hopeless to construct viable models of quintessence. In the following we shall present a number of interesting attempts to obtain quintessence potentials based on particle physics.

7.4.1 A fermion condensate model in global supersymmetric theory

The inverse power-law potential (7.35) appears in globally supersymmetric QCD theories with N_c colors and N_f ($< N_c$) flavors [48]. This theory is based upon quarks ϕ_i ($i = 1, 2, \dots, N_f$) in fundamentals of $SU(N_c)$ and anti-quarks $\bar{\phi}_i$ in anti-fundamentals of $SU(N_c)$. The effective degrees of freedom below the scale of the breaking of the gauge symmetry M are the fermion condensate fields $\Pi_{ij} \equiv \phi_i \bar{\phi}_j$.

From supersymmetry and anomaly-free global symmetries the superpotential W is constrained to be [201]

$$W = (N_c - N_f) M^{\frac{3N_c - N_f}{N_c - N_f}} (\det \Pi)^{-\frac{1}{N_c - N_f}}. \quad (7.55)$$

We are interested in the dynamics of the field expectation values with perturbatively flat directions, i.e. $\langle \phi_{ij} \rangle = \langle \bar{\phi}_{ij}^\dagger \rangle$, where $j = 1, \dots, N_c$ is the gauge index. In this case the so-called D -term, $D^\mu \equiv \phi_i^\dagger t^\mu \phi_i - \bar{\phi}_i t^\mu \bar{\phi}_i^\dagger$, vanishes in the scalar potential (t^μ is a generator of the gauge group). Then, for the flat Kähler potential, the field potential in the global supersymmetric theory is given by

$$V(\phi_i, \bar{\phi}_i) = \sum_{i=1}^{N_f} (|F_i|^2 + |\bar{F}_i|^2), \quad (7.56)$$

where $F_i = \partial W / \partial \phi_i$ and $\bar{F}_i = \partial W / \partial \bar{\phi}_i$. Under gauge and flavor rotations, $\langle \phi_{ij} \rangle$ can be diagonalized in the form $\langle \phi_{ij} \rangle = \langle \bar{\phi}_{ij}^\dagger \rangle = \phi_i \delta_{ij}$ for $1 \leq j \leq N_f$ and $\langle \phi_{ij} \rangle = \langle \bar{\phi}_{ij}^\dagger \rangle = 0$ for $N_f < j \leq N_c$. Taking the expectation values of all N_f scalars to be equal, i.e. $\langle \phi_i \rangle = \phi$ ($i = 1, \dots, N_f$), the determinant in Eq. (7.55) is given by $\det \Pi = \phi^{2N_f}$. Hence the field potential (7.56) reads

$$V(\phi) = \lambda M^{4+n} \phi^{-n}, \quad (7.57)$$

where λ is a dimensionless constant and

$$n = 2 \frac{N_c + N_f}{N_c - N_f}. \quad (7.58)$$

Thus the inverse power-law potential appears with the power n dependent on N_c and N_f . In the tracking regime, Eq. (7.42) gives the field equation of state

$$w_\phi \simeq \frac{N_c + N_f}{2N_c} w_M - \frac{N_c - N_f}{2N_c}. \quad (7.59)$$

From Eq. (7.33) the field ϕ is required to be at least of the order of m_{pl} today. Hence the potential energy needs to satisfy $\lambda M^{4+n} m_{\text{pl}}^{-n} \simeq \rho_c^{(0)}$, where $\rho_c^{(0)}$ is the present cosmological density.

7.4.2 Supergravity models

The fermion condensate model discussed above corresponds to the globally supersymmetric theory. In particular the field ϕ becomes of the order of the Planck mass today. This implies that supergravity corrections to such models need to be taken into account. Recall that the potential for chiral scalar fields φ^i in the presence of supergravity corrections is given by Eq. (6.23). The appearance of the term

$(-3\kappa^2|W|^2)$ can give rise to a negative potential unless the superpotential W and the Kähler potential K are appropriately chosen.

For example, Brax and Martin [50] adopted a superpotential $W = M^{4+\alpha}\varphi^{-\alpha}$ (as in the fermion condensate model) and a flat Kähler potential $K = \varphi\varphi^*$. In this case the kinetic field Lagrangian in Eq. (6.26) yields $\mathcal{L}_{\text{kin}} = -(1/2)\partial_\mu\phi\partial^\mu\phi$ by introducing a real scalar field ϕ defined by $\varphi = \varphi^* = \phi/\sqrt{2}$. Since $D_i W = M^{3+\alpha}\varphi^{-\alpha}(-\alpha\varphi^{-1} + \kappa^2\varphi^*)$ and $K^{\varphi\varphi^*} = 1$ in Eq. (6.23), the field potential is given by

$$V(\phi) = e^{\frac{1}{2}\kappa^2\phi^2} M^{6+2\alpha} \left(\frac{\phi^2}{2}\right)^{-\alpha-1} \left[\alpha^2 - \frac{1}{2}(2\alpha+3)\kappa^2\phi^2 + \frac{1}{4}\kappa^4\phi^4 \right]. \quad (7.60)$$

This potential becomes negative for ϕ of the order of the Planck mass. Hence this cannot be used for realizing the cosmic acceleration today. In order to eliminate the negative contribution, one can impose the constraint $\langle W \rangle = 0$ [50]. In this case the scalar potential is

$$V(\phi) = e^{\kappa^2 K} \left(\frac{\partial W}{\partial \varphi} \right) \left(\frac{\partial W}{\partial \varphi} \right)^* = \frac{\tilde{M}^{6+2\alpha}}{\phi^{2+2\alpha}} e^{\frac{1}{2}\kappa^2\phi^2}, \quad (7.61)$$

where $\tilde{M}^{6+2\alpha} \equiv 2^{\alpha+1}\alpha^2 M^{6+2\alpha}$. This has a potential minimum at $\phi_{\text{min}} = \sqrt{2(1+\alpha)}/\kappa$ with a positive energy $V(\phi) = \tilde{M}^{6+2\alpha}\kappa^{2+2\alpha}(e/(2+2\alpha))^{1+\alpha}$. When $\alpha = 2$, for example, we have that $V(\phi) \simeq 10^{-47} \text{ GeV}^4$ for $\tilde{M} \simeq 10^6 \text{ GeV}$.

While it is possible to realize the condition $\langle W \rangle = 0$ in the presence of matter fields in addition to quintessence [50, 51], this is in general a tight restriction because such a constraint is not easily compatible with models of supersymmetry breaking. Generally, for different choices of the Kähler potential, the field potential can be made positive definite. Recall that the Kähler potential of the form (6.28), which appears at tree-level supersymmetric theories, leads to the cancellation of the negative term $(-3\kappa^2|W|^2)$. Let us then consider a model with the Kähler potential $K = -[\ln(\kappa\varphi + \kappa\varphi^*)]/\kappa^2$ and the superpotential $W = M^{4+\alpha}\varphi^{-\alpha}$ [52]. For this choice the kinetic term becomes canonical by introducing a new scalar field: $\phi = \ln(\kappa\varphi)/(\sqrt{2}\kappa)$. In this case the field potential (6.23) becomes of the exponential type

$$V(\phi) = \tilde{M}^4 e^{-\sqrt{2}\beta\kappa\phi}, \quad (7.62)$$

where $\tilde{M}^4 \equiv \kappa^{\beta+1} M^{\beta+5} (\beta^2 - 3)/2$ and $\beta \equiv 2\alpha + 1$. For the positivity of the potential we require that $\beta > \sqrt{3}$ and hence the slope of the potential, $\lambda \equiv \sqrt{2}\beta$, satisfies the condition $\lambda > \sqrt{6}$. In this case there exists a scaling solution in which the quintessence energy density Ω_ϕ is proportional to the background density Ω_M . Note, however, that the shape of the potential needs to be modified at late times in order to enter the epoch dominated by dark energy.

A more general Kähler potential can also be studied [52]:

$$K = [\ln(\kappa\varphi + \kappa\varphi^*)]^2 / \kappa^2, \quad (7.63)$$

with the superpotential $W = M^{3+\alpha}\varphi^{-\alpha}$. We consider a real scalar field with $\varphi = \varphi^*$. By introducing a new scalar field

$$\phi \equiv \int \sqrt{2K_{\varphi\varphi^*}} d\varphi = -\frac{2}{3\kappa} [1 - \ln(2\kappa\varphi)]^{3/2}, \quad (7.64)$$

the kinetic term becomes canonical [the integration constant is set to be 0 in Eq. (7.64)]. The field potential in this case is given by (see problem 7.4)

$$V = \tilde{M}^4 [2X^2 + (4\alpha - 7)X + 2(\alpha - 1)^2] \frac{1}{X} \exp[(1 - X)^2 - 2\alpha(1 - X)], \quad (7.65)$$

where $\tilde{M}^4 \equiv 2^{2\alpha}\kappa^{2+2\alpha}M^{6+2\alpha}$ and

$$X \equiv 1 - \ln(2\kappa\varphi) = (-3\kappa\phi/2)^{2/3}. \quad (7.66)$$

The field exists in the region $-\infty < \phi < 0$ and hence the parameter X is in the range $0 < X < \infty$. When $|\kappa\phi| \ll 1$ (i.e. $X \ll 1$), we have that $V \propto (-\phi)^{-2/3}$. Meanwhile, when $|\kappa\phi| \gg 1$ (i.e. $X \gg 1$), the potential behaves as $V \propto (-\phi)^{2/3}e^{\phi^{4/3}}$. In the intermediate region there exists a potential minimum at $\phi = \phi_{\min}$ with a positive energy density.

In Fig. 7.6 we plot the field potential for $\alpha = 5$ and $M = 5.36 \times 10^{10}$ GeV. In this case the potential minimum corresponds to $\kappa\phi_{\min} \simeq -0.02$ with $V(\phi_{\min}) \simeq 10^{-47}$ GeV⁴ (i.e. the present cosmological energy density). If the field starts to evolve from the region $|\kappa\phi| \ll 1$, its dynamics during radiation and matter eras is dominated by the potential $V(\phi) \propto \phi^{-2/3}$. Then the field will show the tracking behavior as we have already explained previously for the inverse power-law potential. Since the potential minimum exists in the present case, the field is eventually trapped at this point so that the equation of state approaches $w_\phi = -1$. If the field is initially in the region $|\kappa\phi| \gg 1$, the exponential term in Eq. (7.65) is important so that the quintessence density exhibits a scaling-like behavior relative to the background density [52]. The field finally approaches the potential minimum and leads to the late-time cosmic acceleration.

A problem for supersymmetric quintessence models is associated with the fact that supersymmetry must be broken if it is to be realized at all in nature. The supersymmetry breaking is supposed to occur for the scale $M_S \gtrsim 10^{10}$ GeV in the gravity-mediated scenario to have a viable low-energy phenomenology. The squared of the breaking scale M_S is related to the first term (so-called F -term) on the r.h.s. of Eq. (6.23), i.e. $\langle F \rangle \sim M_S^2$, if we write the potential (6.23) in the form $V = F^2 - 3e^{\kappa^2 K} |W|^2$. In order to cancel the F -term contribution and give only a small

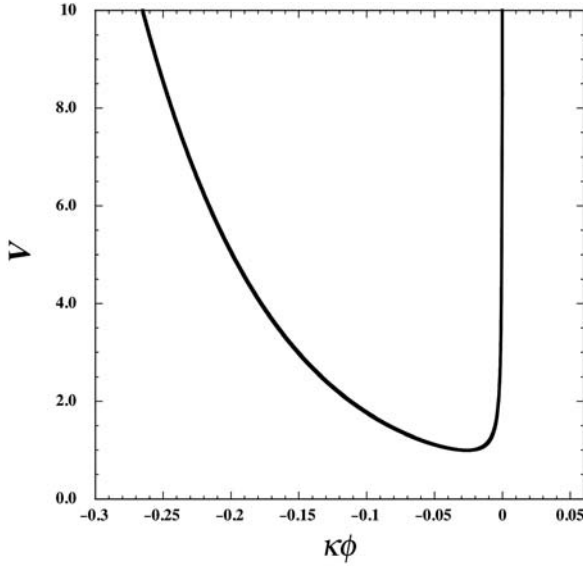


Figure 7.6 The potential (7.65) in the unit of GeV^4 (multiplied by the factor 10^{47}) for the model parameters $\alpha = 5$ and $M = 5.36 \times 10^{10} \text{ GeV}$. There exists a potential minimum around $\kappa\phi = -0.02$ in this case.

amount of dark energy we require that $W \sim \langle F \rangle \kappa^{-1} \sim m_{3/2} \kappa^{-2}$, where $m_{3/2}$ is the gravitino mass.² Recall that the superpotential $W = M^{3+\alpha} \varphi^{-\alpha}$ adopted to derive the potential (7.65) corresponds to the scale $W \sim M^{3+\alpha} \kappa^\alpha \sim (10^{-3} \text{ eV})^2 \kappa^{-1} \ll \langle F \rangle \kappa^{-1}$ today, where we have used the fact that $\tilde{M}^4 \sim (M^{3+\alpha} \kappa^{1+\alpha})^2 \sim (10^{-3} \text{ eV})^4$. This shows that the superpotential $W = M^{3+\alpha} \varphi^{-\alpha}$ cannot be a dominant source for the supersymmetry breaking. If we consider the superpotential of the form $W = M^{3+\alpha} \varphi^{-\alpha} + m_{3/2} \kappa^{-2}$, then the scalar potential acquires the following modification [52]

$$\delta V = m_{3/2}^2 \kappa^{-2} + m_{3/2} M^{3+\alpha} \kappa^{-\alpha}. \quad (7.67)$$

The first term leads to a significant disruption of the quintessence potential. The supersymmetry breaking with $\langle F \rangle > 10^{10} \text{ GeV}$ gives rise to an energy scale of the order $m_{3/2}^2 / M_{\text{pl}}^2$, which is much greater than the scale of dark energy.

While the above problem looks serious, it is anticipated that some string theory models may allow a possibility to overcome this problem due to mechanisms of unconventional supersymmetry breaking. For example, Witten [202] advocated that we may live in a 4-dimensional world with unbroken supersymmetry and that

² The gravitino is a supersymmetric fermionic partner of the graviton. Since it mediates supergravity interactions, the broken supersymmetry gives rise to a gravitino mass directly related to the supersymmetry breaking scale.

the mass splitting between the superpartners occurs as a result of the excitations of the system while maintaining a supersymmetric ground state. If this is the case, we do not need to worry about the contribution of the supersymmetry breaking terms to the quintessence potential. Recall also that in “no-scale” supersymmetric models discussed in Section 6.5.1 the supersymmetry breaking F -term vanishes at tree-level. Hence there exists “no-scale supersymmetry” to protect a small mass at least in the framework of supergravity.

We also have a class of supergravity models where the above-mentioned problem can be avoided. In the so-called $\mathcal{N} \geq 2$ extended supergravity models the mass squared of any ultra-light scalar fields can be quantized in unit of squared of the Hubble constant H_0 of de Sitter solutions [203, 204, 205]. The de Sitter solutions correspond to the extrema of an effective potential $V(\phi)$ for some scalar field ϕ . Around the extremum at $\phi = 0$ the field potential is given by $V(\phi) = \Lambda + (1/2)m_\phi^2\phi^2$ with $\Lambda > 0$. In extended supergravity theories the mass m_ϕ is always related to Λ via the relation $m_\phi^2 = n\Lambda/(3M_{\text{pl}}^2)$, where n is an integer of the order of unity [203, 204]. Since $H_0^2 = \Lambda/(3M_{\text{pl}}^2)$ in the de Sitter space, it follows that $m_\phi^2 = nH_0^2$. In the $\mathcal{N} = 2$ gauged supergravity we have $n = 6$ for a stable de Sitter vacuum [205], giving the field potential

$$V(\phi) = 3H_0^2 M_{\text{pl}}^2 \left[1 + (\phi/M_{\text{pl}})^2 \right]. \quad (7.68)$$

The $\mathcal{N} = 8$ supergravity theories correspond to a negative mass squared $m_\phi^2 = -6H_0^2$ [204] and the resulting potential is

$$V(\phi) = 3H_0^2 M_{\text{pl}}^2 \left[1 - (\phi/M_{\text{pl}})^2 \right]. \quad (7.69)$$

The constant Λ determines the energy scale of the supersymmetry breaking. In order for the above models to be compatible with the cosmic acceleration today we require that $\Lambda \simeq H_0^2 M_{\text{pl}}^2 \simeq 10^{-47} \text{ GeV}^4$. The supersymmetry breaking scale is so small that the ultra-light mass m_ϕ of the order of 10^{-33} eV can be protected against quantum corrections. Note that the potentials (7.68) and (7.69) fall into the class of thawing models in which the field is nearly frozen until recently and starts to evolve when the Hubble parameter decreases and becomes comparable to $|m_\phi|$.

7.4.3 Pseudo-Nambu–Goldstone models

There is another class of models based on Pseudo-Nambu–Goldstone Bosons (PNGB) to protect the light mass of quintessence by the $U(1)$ symmetry. An example of a very light PNGB is the axion field, which is associated with the Peccei–Quinn (PQ) symmetry [79] introduced to address the strong CP problem.

When a global $U(1)_{\text{PQ}}$ symmetry is spontaneously broken, the axion appears as an angular massless field ϕ with an expectation value $\langle\phi\rangle = f_s e^{i\phi/f_s}$ of a complex scalar at the scale f_s .

The key model parameters of PNGB models are the mass scale of spontaneous symmetry breaking f and the energy scale of explicit symmetry breaking μ . The representative PNGB potential is given by [47]

$$V(\phi) = \mu^4 [1 + \cos(\phi/f)] . \quad (7.70)$$

The mass squared of the field, $m_\phi^2 = V_{,\phi\phi}$, is $m_\phi^2 = -\mu^4/f^2$ at $\phi = 0$. In the limit that $\mu \rightarrow 0$ the potential (7.70) vanishes so that the symmetry becomes exact. In this situation radiative corrections to V do not give rise to an explicit symmetry-breaking term because they are simply proportional to μ^4 . Hence the small mass associated with dark energy ($m_\phi \simeq 10^{-33}$ eV) can be protected against radiative corrections [47].

If the PNGB potential (7.70) is responsible for the cosmic acceleration today, we require that $H_0^2 \approx \mu^4/M_{\text{pl}}^2$. Hence the field mass squared around $\phi = 0$ can be estimated as

$$m_\phi^2 \approx -\frac{M_{\text{pl}}^2}{f^2} H_0^2 . \quad (7.71)$$

The slow-roll condition, $|\eta_s| = |V_{,\phi\phi}/(\kappa^2 V)| \lesssim 1$, leads to $f \gtrsim M_{\text{pl}}$. From Eq. (7.71) the field mass is constrained to be $|m_\phi| \lesssim H_0$. Hence the field begins to evolve only recently from the potential maximum when the Hubble rate H drops down to the order of $|m_\phi|$. The PNGB model, which belongs to the class of thawing quintessence models, leads to similar cosmological evolution to the corresponding evolution in the $\mathcal{N} = 8$ supergravity model in Eq. (7.69).

There have been a number of attempts to explain the small energy scale $\mu \approx 10^{-3}$ eV required for the PNGB quintessence in supersymmetric theories [206, 207, 208, 209]. For instance, Hall *et al.* [209] proposed an interesting idea to relate μ with two fundamental scales, the Planck scale $M_{\text{pl}} \approx 10^{18}$ GeV and the electroweak scale $v \approx 10^3$ GeV. There is an induced seesaw scale, $v^2/M_{\text{pl}} \approx 10^{-3}$ eV, which is the same order as μ . If we assume the relation $\mu \approx v^2/M_{\text{pl}}$ then we have $|m_\phi^2| \approx \mu^4/f^2 \approx v^8/M_{\text{pl}}^6$, where we have used $f \approx M_{\text{pl}}$ (f can be fixed because M_{pl} and v are two fundamental scales of the theory). This gives rise to the mass of the order

$$|m_\phi| \approx \frac{v^4}{M_{\text{pl}}^3} \approx 10^{-33} \text{ eV} , \quad (7.72)$$

as required for quintessence.

How is the relation $\mu \approx v^2/M_{\text{pl}}$ justified? Hall *et al.* considered models with an axion in a hidden sector of the theory [209]. If the fundamental scale of supersymmetry breaking in nature is of the order of the TeV scale (i.e. $v \approx 10^3 \text{ GeV}$), any sector of the theory that feels this symmetry breaking indirectly through gravity mediation has an effective supersymmetry breaking scale $m_B = v^2/M_{\text{pl}}$. Let us consider the case in which the hidden sector has a supersymmetric gauge interaction which acts on chiral superfields Q and Q^c . After the supersymmetry breaking the squarks and gluinos acquire a mass of the order of m_B . We also assume that the quark acquires the same order of mass. If the hidden sector has a PQ symmetry broken at f of the order of M_{pl} , the interacting Lagrangian between the axion ϕ and the quarks q, q^c at the scale M is given by

$$\mathcal{L}_{\text{axion}} = m_q q q^c e^{i\phi/f} + \text{h.c.}, \quad (7.73)$$

where $m_q (\approx m_B)$ is the quark mass. The axion has a PQ symmetry $U(1)_{\text{PQ}}$ broken around the Planck scale, whereas the quark bilinear $q q^c$ has an axial $U(1)$ symmetry denoted as $U(1)_A$. The interaction (7.73) breaks the $U(1)_{\text{PQ}} \times U(1)_A$ symmetry explicitly. If at least one of the quark flavors in Eq. (7.73) has a mass smaller than the order of M then a quark condensate forms such that $\langle q q^c \rangle \approx M^3 e^{i\tilde{\phi}/M}$ with an angular field $\tilde{\phi}$. Then the interaction term (7.73) gives rise to the following potential

$$V = \mu^4 \cos\left(\frac{\phi}{f} + \frac{\tilde{\phi}}{M}\right), \quad \mu^4 = m_q M^3. \quad (7.74)$$

Since both M and m_q are close to m_B , it follows that the scale μ is of the order of the seesaw scale m_B , i.e. $\mu \approx m_B = v^2/M_{\text{pl}}$. Interestingly the axion model in a hidden sector provides a natural explanation for the smallness of μ as well as m_ϕ without having an instability problem against radiative corrections.

7.4.4 Dilatonic quintessence

In string theory, gauge and gravitational couplings are not fixed a priori, but are related to the vacuum expectation value of a scalar field, the dilaton ϕ [64, 210]. Precisely speaking, at the tree level in the string loop expansion, the vacuum expectation value of the 4-dimensional dilaton ϕ is related to the gauge coupling α_{GUT} and to the (reduced) Planck mass M_{pl} through $e^\phi \simeq M_s^2/M_{\text{pl}}^2 \simeq \alpha_{\text{GUT}}$, where $M_s \equiv \sqrt{2/\alpha'}$ is the string mass and α' is a universal Regge slope parameter of the string. The weak coupling regime corresponds to the dilaton coupling with $e^\phi \ll 1$.

The presence of the dilatonic coupling can give rise to the violation of the equivalence principle as well as the variation of the coupling constants. Usually it

is assumed that the dilaton and other moduli of string theory acquire masses by some non-perturbative mechanism, in such a way that their long-range interactions are suppressed and their vacuum expectation values (VEVs) are frozen at some phenomenologically reasonable value. An alternative approach is that the dilaton decouples from other fields at the level of the low-energy effective action in the regime $e^\phi \gtrsim 1$. This is the runaway dilaton scenario [211, 212] in which the gauge and gravitational coupling with dilaton approaches constant values in the limit that $e^\phi \rightarrow \infty$.

To be more explicit, let us consider a generic effective string action at lowest order in the Regge slope parameter α' :

$$S = \frac{1}{\alpha'} \int d^4x \sqrt{-g} \left[B_g(\phi) R + B_\phi(\phi) g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - \alpha' U(\phi) \right. \\ \left. + \text{higher order terms} \right] + S_m[\phi, g_{\mu\nu}, \Psi_i], \quad (7.75)$$

where $U(\phi)$ is a potential of the dilaton field ϕ and S_m is the action for matter fields Ψ_i which are generally coupled to the dilaton. The dilaton-dependent loop effects as well as the non-perturbative corrections are encoded, at the lowest order of approximation in α' , in the coupling functions $B_i(\phi)$ (where $i = g, \phi, \dots$). In the weak coupling regime where the tree-level string coupling $g_s^2 \equiv e^\phi$ is much smaller than unity, the functions $B_i(\phi)$ can be expressed by an expansion of the form

$$B_i(\phi) = e^{-\phi} + c_0 + c_1 e^\phi + \dots \quad (e^\phi \ll 1). \quad (7.76)$$

In the weak coupling limit ($g_s^2 \rightarrow 0$) the first term dominates to give the couplings $B_i(\phi) \simeq e^{-\phi}$.

In the runaway dilaton scenario the effective couplings $B_i(\phi)$ reach extrema at infinite coupling $e^\phi \rightarrow \infty$ and they exhibit the following general behavior

$$B_i(\phi) = C_i + D_i e^{-\phi} + \mathcal{O}(e^{-2\phi}) \quad (e^\phi \gg 1). \quad (7.77)$$

This is motivated by the fact that the couplings B_i are, for the most part, induced by the quantum corrections of many moduli and gauge bosons of the theory [213]. In this scenario the dilaton gradually decouples from gravity and the matter fields by evolving towards infinity: $\phi \rightarrow \infty$. The coefficients C_i and D_i are of order 10^2 and unity, respectively, where the coefficients C_i correspond to the number of independent degrees of freedom which have been integrated over.

The action (7.75) can be transformed to the so-called Einstein frame action in which the Ricci scalar $\tilde{R}$ does not have an explicit coupling with the field ϕ . We introduce the conformally related Einstein frame metric

$$\tilde{g}_{\mu\nu} = \Omega^2 g_{\mu\nu}, \quad (7.78)$$

where Ω^2 is called a conformal factor. For the action of the form $S = \int d^4x \sqrt{-g} f(\phi, R) + \dots$ we obtain the Einstein frame action $S_E = \int d^4x \sqrt{-\tilde{g}} \tilde{R}/(2\kappa^2) + \dots$ by choosing [214]³

$$\Omega^2 = 2\kappa^2 \frac{\partial f}{\partial R}. \quad (7.79)$$

For the theories given by (7.75), this corresponds to

$$\Omega^2 = 2\kappa^2 \frac{B_g}{\alpha'} = B_g \frac{M_s^2}{M_{\text{pl}}^2}, \quad (7.80)$$

where we have introduced the string mass scale $M_s = \sqrt{2/\alpha'}$ as well as the reduced Planck mass $M_{\text{pl}} = 1/\kappa$. Since the string frame action (7.75) approaches the General Relativistic action in the limit $\phi \rightarrow \infty$, it follows that $M_{\text{pl}}^2 \simeq C_g M_s^2$ [211, 212]. Under the conformal transformation (7.78) with (7.80), the action in the Einstein frame reads

$$S_E = \int d^4x \sqrt{-\tilde{g}} \left[\frac{M_{\text{pl}}^2}{2} \tilde{R} - \frac{\epsilon}{2} \tilde{g}^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - V(\phi) + \text{higher order terms} \right] + S_m[\phi, \Omega^{-2} \tilde{g}_{\mu\nu}, \Psi_i], \quad (7.81)$$

where we have introduced a canonically defined scalar field ϕ with the dimensions of mass and the Einstein frame potential $V(\phi)$ as follows

$$M_{\text{pl}}^2 \left[\frac{3}{2} \left(\frac{1}{B_g} \frac{dB_g}{d\phi} \right)^2 - \frac{B_\phi}{B_g} \right] d\phi^2 = \epsilon d\phi^2, \quad V(\phi) = U(\phi)/\Omega^4, \quad (7.82)$$

with $\epsilon = \pm 1$. If the expression in the square bracket of Eq. (7.82) is positive, then the dilaton behaves as a normal (non-ghost) scalar field ($\epsilon = +1$). For the tree-level action in the weak coupling limit ($e^\phi \rightarrow 0$), we have $B_\phi = B_g = e^{-\phi}$ so that the sign of the field kinetic energy is normal in the Einstein frame. There is the possibility that ϵ becomes negative depending on the coupling functions $B_g(\phi)$ and $B_\phi(\phi)$. Although such a ghost scalar field is generally plagued by an instability problem of quantum fluctuations, it is possible to make the system stable by taking into account higher-order field derivative corrections to the action (7.75). We shall discuss this possibility in the next chapter in connection with k-essence.

Since the potential $U(\phi)$ for the dilaton vanishes at the tree level, it must be of non-perturbative origin with an exponential damping in the weak coupling limit ($\phi \rightarrow -\infty$). We shall assume that it also goes to zero in the regime $\phi \rightarrow \infty$. A viable ansatz is [211]

$$U(\phi) = M_0^4 [\exp(-e^{-\phi}/\beta_1) - \exp(-e^{-\phi}/\beta_2)], \quad (7.83)$$

³ See also Chapter 9 for detailed explanation about the conformal transformation.

where M_0 is some mass scale, and β_1 and β_2 are constants satisfying $0 < \beta_2 < \beta_1$. This is a bell-type potential having a maximum in an intermediate regime between $\phi \rightarrow -\infty$ and $\phi \rightarrow \infty$. In the regime $e^\phi \gg 1$ the potential behaves as

$$U(\phi) \simeq \left(\frac{1}{\beta_2} - \frac{1}{\beta_1} \right) M_0^4 e^{-\phi}. \quad (7.84)$$

We need to caution that the matter fields Ψ_i have direct couplings with the dilaton field in the Einstein frame. In this sense, when $\epsilon = +1$, the Einstein frame action (7.81) corresponds to a coupled quintessence scenario [17] with an exponential potential (7.84) in the regime $e^\phi \gg 1$. If this coupling is strong, it can give rise to a violation of equivalence principle as well as a non-standard matter-dominated epoch inconsistent with observations. The strength of the matter coupling $Q(\varphi)$ is proportional to the derivative $\delta S_m / \delta \varphi$. In the runaway dilaton scenario the matter fields gradually decouple from the field φ so that the coupling decreases to $Q(\varphi) \rightarrow 0$ in the limit $e^\phi \gg 1$. In Ref. [211] it was shown that the runaway dilaton can be responsible for the present cosmic acceleration while satisfying the equivalence principle constraints. In Chapter 8 we shall present the detailed cosmological dynamics of the coupled quintessence scenario and will study the consistency with local experimental constraints.

7.5 Reconstruction of quintessence from observations

From the point of view of cosmological observations, the crucial differences between the cosmological constant and quintessence are that the equation of state of the latter changes in time and that quintessence fluctuates and, at some level, may cluster. Here we discuss the possibility to distinguish quintessence models from the Λ CDM model from the background cosmic expansion history. In Chapter 11 we discuss the clustering properties of dark energy.

7.5.1 Reconstructing the potential and the equation of state of quintessence

In SN Ia observations the Hubble parameter $H(z)$ is estimated by measuring the luminosity distance $d_L(z)$, see Eq. (2.86). This allows for the reconstruction of the equation of state of dark energy $w_{\text{DE}}(z)$ by using the relation (2.88). In the case of quintessence it is also possible to reconstruct the field potential $V(\phi)$. Let us consider a non-relativistic matter with an equation of state $w_m = 0$ as well as a quintessence field ϕ . For the analysis of SN Ia observations we can neglect radiation. Using the relations $\rho_m = \rho_m^{(0)}(1+z)^3$ and $dt = -dz/[H(1+z)]$, we find that Eqs. (7.7) and (7.8) can be rewritten as [215, 216, 217, 218, 219]

(see problem 7.5)

$$\frac{\kappa^2}{2} \left(\frac{d\phi}{dz} \right)^2 = \frac{1}{1+z} \frac{d \ln E(z)}{dz} - \frac{3\Omega_m^{(0)}}{2} \frac{1+z}{E^2(z)}, \quad (7.85)$$

$$\frac{\kappa^2 V}{3H_0^2} = E(z) - \frac{1+z}{6} \frac{dE^2(z)}{dz} - \frac{1}{2} \Omega_m^{(0)} (1+z)^3. \quad (7.86)$$

Since $(d\phi/dz)^2$ is positive, it follows from Eq. (7.85) that the condition,

$$\frac{dH^2}{dz} \geq 3\Omega_m^{(0)} H_0^2 (1+z)^2, \quad (7.87)$$

must be satisfied. This corresponds to the weak energy condition, $\rho_\phi + P_\phi \geq 0$. The field ϕ is known as a function of z by integrating Eq. (7.85). Inverting $\phi(z)$ to $z(\phi)$ and substituting $z(\phi)$ into Eq. (7.86), it is possible to reconstruct the potential V with respect to ϕ by using the information of the observationally known values of $H(z)$ and $H'(z)$ as well as $\Omega_m^{(0)}$. Independent information on $\Omega_m^{(0)}$ has to be obtained from large-scale structure.

The reconstruction process is however subject to two general problems. The first is that finding a model containing a trajectory with a given expansion rate does not guarantee that the trajectory is stable. It is possible in fact that even if a reconstructed solution exists it is never reached by realistic initial conditions or that any arbitrarily small perturbation from the trajectory itself leads far from the solution. After reconstructing a particular model one should then explicitly check its stability.

The second problem is that the actual observational data such as the luminosity distance $d_L(z)$ are known at discrete values of redshifts. Moreover the data are affected by systematic and statistical errors. Hence it is not possible to directly differentiate $d_L(z)$ with respect to z in Eq. (2.86) to obtain $H(z)$. We require some smoothing process for reconstructing the equation of state $w_{DE}(z)$ and the quintessence potential $V(\phi)$. In order to address such a smoothing problem several authors have assumed parametric forms of $d_L(z)$ [215, 216], or $H(z)$ [219, 220, 221], or $w_{DE}(z)$ [222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234]. For example, if we parametrize $w_{DE}(z)$ as a function of z , the Hubble parameter in the flat Universe is known to be [see Eq. (2.84)]

$$H^2(z) = H_0^2 \left[\Omega_m^{(0)} (1+z)^3 + \Omega_{DE}^{(0)} \exp \left\{ \int_0^z \frac{3(1+w_{DE}(\tilde{z}))}{1+\tilde{z}} d\tilde{z} \right\} \right]. \quad (7.88)$$

Then the integration of Eq. (2.87) gives the luminosity distance $d_L(z)$ as a function of z , which we can confront with observations.

Several parametrizations of $w_{\text{DE}}(z)$ have been proposed so far. We can write such parametrizations in the form

$$w_{\text{DE}}(z) = \sum_{n=0} w_n x_n(z), \quad (7.89)$$

where the expansions can be given by

$$(i) \text{ Redshift: } x_n(z) = z^n, \quad (7.90)$$

$$(ii) \text{ Scale factor: } x_n(z) = \left(1 - \frac{a}{a_0}\right)^n = \left(\frac{z}{1+z}\right)^n, \quad (7.91)$$

$$(iii) \text{ Logarithmic: } x_n(z) = [\ln(1+z)]^n. \quad (7.92)$$

Parametrization (i) was introduced by Huterer and Turner [216] and Weller and Albrecht [224] with $n \leq 1$, i.e. $w_{\text{DE}} = w_0 + w_1 z$. Parametrization (ii) with $n \leq 1$ was proposed by Chevalier and Polarski [223] and Linder [228]:

$$w_{\text{DE}}(z) = w_0 + w_1(1 - a) = w_0 + w_1 \frac{z}{1+z}, \quad (7.93)$$

which behaves as $w_{\text{DE}}(z) = w_0 + w_1$ for $z \rightarrow \infty$ and $w_{\text{DE}}(z) \rightarrow w_0$ for $z \rightarrow 0$. Jassal *et al.* [231] extended this to a more general form with $w_{\text{DE}}(z) = w_0 + w_1 z/(1+z)^p$. Parametrization (iii) with $n \leq 1$ was introduced by Efstathiou [222]. While the Taylor expansions are taken at linear order ($n \leq 1$) for the above parametrizations, one can also adopt functional forms that can be applied to the case of a fast transition of $w_{\text{DE}}(z)$ (e.g., [226], [227]).

Figure 7.7 shows the observational allowed evolution of $w_{\text{DE}}(z)$ for the parametrization (7.93). This is based on the constraint coming from SN Ia gold data combined with WMAP 3-year data and SDSS data. The time-varying equation of state of dark energy such as quintessence is confined in the region $-1 < w_{\text{DE}}(z) \lesssim -0.8$. While the quintessence is restricted to be in the parameter region $w_{\text{DE}}(z) \geq -1$, the observational data allow the phantom equation of state $w_{\text{DE}}(z) < -1$. In Chapter 9 we will see that it is possible to realize $w_{\text{DE}}(z) < -1$ in modified gravity theories.

The quintessence potential (7.86) can be also reconstructed from observations by parametrizing $w_{\text{DE}}(z)$. As long as the potential is a smooth and slowly varying function with respect to ϕ , the equation of state $w_{\text{DE}}(z)$ should not change rapidly. For quintessence models in which $w_{\text{DE}}(z)$ evolves moderately with the redshift, the two-parameter fit (7.93) can approximately reproduce the quintessence dynamics in most cases. However, if the quintessence potential has abrupt features, the equation of state can change rapidly. In such cases the parametrization proposed in Refs. [226, 227] is more suitable than the two-parameter forms given above.

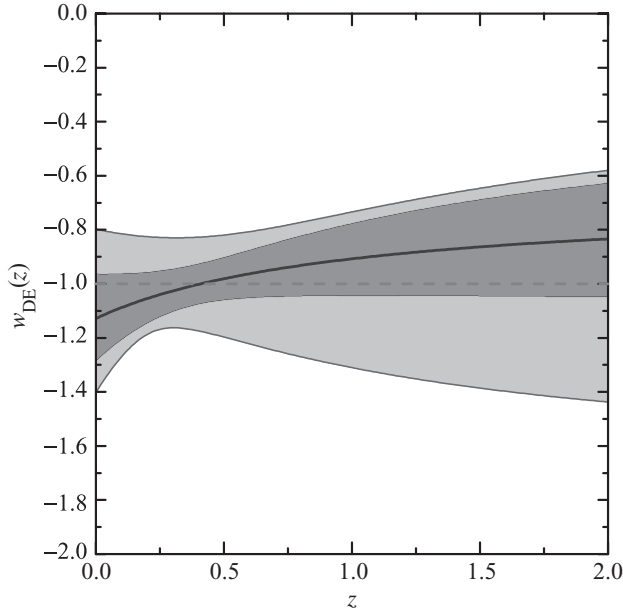


Figure 7.7 Evolution of the equation of state of dark energy reconstructed from the 157 “gold” SN Ia data, WMAP 3-year data, and SDSS. The central line, the dark grey region, and light grey region represent the best-fit, 1σ region, and 2σ region, respectively. The parametrization, $w_{\text{DE}}(z) = w_0 + w_1 z/(1+z)$, is used for the data analysis. From Ref. [235].

It is clear that these parametrizations are purely phenomenological. They are however a necessary step towards a more complete characterization of dark energy and are routinely employed to analyze data, to optimize survey design and to compare results. As such, they are an invaluable tool, but one has to remember that in most cases the results will depend on the chosen parametrization.

7.5.2 Statefinder

From the SN Ia and other observations, we know that the sign of the deceleration parameter,

$$q \equiv -\frac{\ddot{a}}{aH^2}, \quad (7.94)$$

has changed from positive to negative around the redshift $z = \mathcal{O}(1)$ (although in Section 8.3.2 we will see an alternative interpretation). In future high-precision observations it may be possible to obtain information on the variation of q . In this regard, Sahni *et al.* [236] introduced the following “statefinder” parameters

$$r \equiv \frac{\ddot{a}}{aH^3}, \quad s \equiv \frac{r-1}{3(q-1/2)}. \quad (7.95)$$

These two statefinders allow us to distinguish quintessence and other dark energy models from the Λ CDM model. Note that the definition in Eq. (7.95) is based on geometrical quantities. Hence the statefinders can be applied to modified gravity models as well.

Let us consider the Einstein gravity in a Universe dominated by dark energy and non-relativistic fluid. The evolution equations are

$$3H^2 = \kappa^2 (\rho_{\text{DE}} + \rho_m) , \quad (7.96)$$

$$2\dot{H} = -\kappa^2 [(1 + w_{\text{DE}})\rho_{\text{DE}} + \rho_m] , \quad (7.97)$$

where ρ_{DE} and ρ_m satisfy the continuity equations, $\dot{\rho}_{\text{DE}} = -3H(1 + w_{\text{DE}})\rho_{\text{DE}}$ and $\dot{\rho}_m = -3H\rho_m$, respectively. The first statefinder is given by $r = 1 + 3\dot{H}/H^2 + \ddot{H}/H^3$ and hence it is calculated by taking the time-derivative of Eq. (7.97). Using Eq. (7.97) we find that the deceleration parameter can be expressed as $q = 1/2 + (3/2)w_{\text{DE}}\Omega_{\text{DE}}$, where $\Omega_{\text{DE}} \equiv \kappa^2 \rho_{\text{DE}}/(3H^2)$. Then the statefinders in the Einstein gravity read

$$r = 1 + \frac{9w_{\text{DE}}\Omega_{\text{DE}}}{2}s , \quad (7.98)$$

$$s = 1 + w_{\text{DE}} - \frac{\dot{w}_{\text{DE}}}{3w_{\text{DE}}H} . \quad (7.99)$$

We note that the second statefinder can be also expressed as

$$s = \frac{1 + w_{\text{DE}}}{w_{\text{DE}}} \frac{\dot{P}_{\text{DE}}}{\dot{\rho}_{\text{DE}}} , \quad (7.100)$$

where $P_{\text{DE}} = w_{\text{DE}}\rho_{\text{DE}}$. The Λ CDM model ($w_{\text{DE}} = -1$) corresponds to the fixed point $(r, s) = (1, 0)$. For general dark energy models the cosmological trajectories evolve in the (r, s) plane, so that they can be distinguished from the Λ CDM model.

Let us consider quintessence models (with w_{DE} replaced by w_ϕ). The field potential giving a constant equation of state w_ϕ ($-1 < w_\phi < 0$) is [3, 237]

$$V(\phi) = \frac{3H_0^2(1 - w_\phi)(1 - \Omega_m^{(0)})^{1/|w_\phi|}}{2\kappa^2\Omega_m^{(0)\beta}} \times \sinh^{-2\beta} \left(|w_\phi| \sqrt{\frac{3\kappa^2}{4(1 + w_\phi)}} (\phi - \phi_0 + \phi_1) \right) , \quad (7.101)$$

where ϕ_0 is the field value today and

$$\beta \equiv \frac{1 + w_\phi}{|w_\phi|} , \quad \phi_1 \equiv \sqrt{\frac{4(1 + w_\phi)}{3\kappa^2}} \frac{1}{|w_\phi|} \ln \left(\frac{1 + \sqrt{1 - \Omega_m^{(0)}}}{\sqrt{\Omega_m^{(0)}}} \right) . \quad (7.102)$$

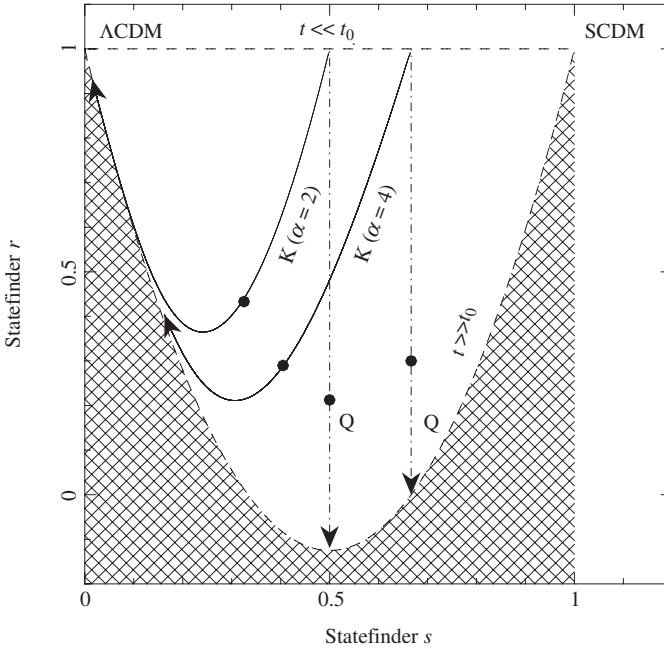


Figure 7.8 The statefinder pair (r, s) for the constant w_ϕ model and the power-law quintessence model with the potential $V(\phi) = M^{4+\alpha}\phi^{-\alpha}$. The Λ CDM model ($w_{\text{DE}} = -1$) corresponds to $(r, s) = (1, 0)$ while SCDM denotes standard CDM (i.e. flat space, no dark energy) with $(r, s) = (1, 1)$. For the constant w_ϕ model, s remains fixed at $s = 1 + w_\phi$ while r asymptotically decreases to $r \simeq 1 + 9w_\phi(1 + w_\phi)/2$. The two cases with $w_\phi = -1/3, -1/2$ are shown in the figure. The power-law quintessence model starts to evolve from the tracking regime with $w_\phi \simeq -2/(\alpha + 2)$ and $\Omega_\phi \ll 1$, which corresponds to $r \rightarrow 1$ and $s \rightarrow \alpha/(\alpha + 2)$. In this case the solutions approach $(r, s) = (1, 0)$ at late times. The hatched region is the forbidden region for these models. The filled circles represent the values of (r, s) today ($\Omega_m^{(0)} \simeq 0.3$). From Ref. [236].

The potential (7.101) can be derived by integrating Eq. (2.88) for constant $w_{\text{DE}} = w_\phi$ to get $H(z)$ and then by using Eqs. (7.85) and (7.86). For constant w_ϕ the parameter s is constant ($s = 1 + w_\phi$), while the parameter r decreases from $r = 1$ to $r = 1 + 9w_\phi(1 + w_\phi)/2$ as the dark energy density grows from $\Omega_\phi = 0$ to $\Omega_\phi = 1$. In the (r, s) plane the trajectory in this case is characterized by a vertical line, see Fig. 7.8.

Let us consider the power-law potential $V(\phi) = M^{4+\alpha}\phi^{-\alpha}$ as an example of the freezing quintessence models. The solutions are in the tracking regime during most of the matter era, so that the equation of state of quintessence is given by $w_\phi \simeq -2/(\alpha + 2)$, see Eq. (7.42). Since $\Omega_\phi \ll 1$ in the early epoch, we have that $r \simeq 1$ and $s \simeq \alpha/(\alpha + 2)$ at the beginning. The solutions asymptotically approach the de Sitter point $(r, s) = (1, 0)$ by following the curve $r = 1 + 9s(s - 1)/2$ from

some time onward (note that this curve corresponds to the limit $\Omega_\phi \rightarrow 1$ for the constant w_ϕ case). The cosmological evolution is confined in the region $s > 0$ in this case, which translates into the condition

$$\dot{w}_\phi/H > 3w_\phi(1 + w_\phi), \quad (7.103)$$

where we have used the fact that $w_\phi < 0$.

In the thawing models of quintessence such as the PNGB model the field is initially in the region $w_\phi \simeq -1$ and $\Omega_\phi \ll 1$ so that the trajectories start from the de Sitter point $(r, s) \simeq (1, 0)$. Since $\dot{w}_\phi > 0$ and $-1 < w_\phi < 0$, Eq. (7.99) shows that $s > 0$ and hence $r < 1$ from Eq. (7.98). Then the cosmological trajectories evolve to the opposite direction compared to the freezing models. Note that the allowed parameter space in the (r, s) plane is confined in the region $r < 1$ and $s > 0$ in both freezing and thawing models.

The parameter spaces in the (r, s) plane are different depending on the models of dark energy. Let us consider for instance the Chaplygin gas model (we will discuss it fully in Section 8.6) in which the pressure P_{DE} is related to the energy density ρ_{DE} via $P_{\text{DE}} = -A/\rho_{\text{DE}}$ ($A > 0$) [45]. It follows from Eq. (7.100) that the second statefinder in this case is given by

$$s = \frac{1 + w_{\text{DE}}}{w_{\text{DE}}} \frac{A}{\rho_{\text{DE}}^2}. \quad (7.104)$$

Since $-1 < w_{\text{DE}} < 0$ we have that $s < 0$ and hence $r > 1$ from Eq. (7.98). In this way the Chaplygin gas model can be distinguished from the quintessence model.

7.5.3 Quintessence in the (w_ϕ, w'_ϕ) plane

The statefinder parameters (r, s) involve the time derivative of w_{DE} in their expressions. In the case of quintessence the sign of $\dot{w}_\phi$ is important to distinguish the freezing models from the thawing models. The freezing models and the thawing models are characterized by the conditions $w'_\phi \equiv dw_\phi/dN < 0$ and $w'_\phi > 0$, respectively. Let us consider the allowed parameter region of both models in the (w_ϕ, w'_ϕ) plane.

First of all, the freezing models satisfy the condition (7.103), i.e. $w'_\phi > 3w_\phi(1 + w_\phi)$. According to the analysis of a variety of freezing quintessence models there is an upper bound for the variation of the equation of state: $w'_\phi \lesssim 0.2w_\phi(1 + w_\phi)$ [182]. Note that this bound comes from the evolution since $z \sim 1$. Thus the allowed region for the freezing models is given by

$$3w_\phi(1 + w_\phi) < w'_\phi \lesssim 0.2w_\phi(1 + w_\phi), \quad (7.105)$$

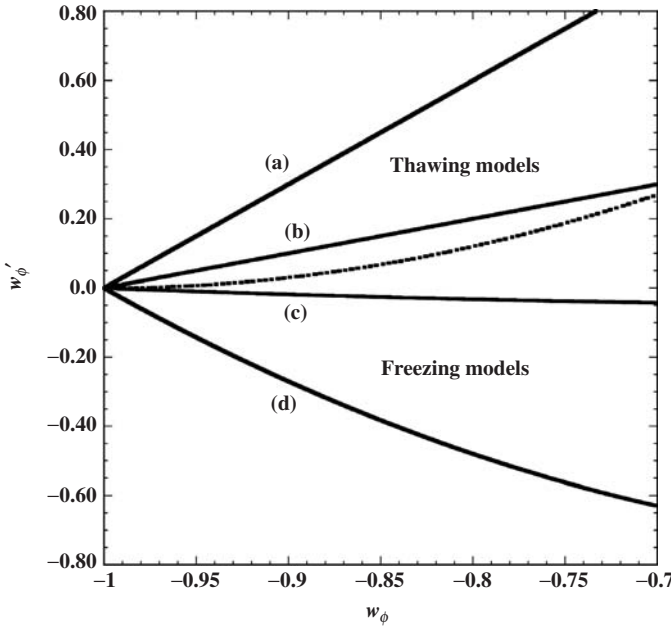


Figure 7.9 The allowed region in the (w_ϕ, w'_ϕ) plane for thawing and freezing models of quintessence (here primes denote d/dN). The thawing models correspond to the region between two curves: (a) $w'_\phi = 3(1 + w_\phi)$ and (b) $w'_\phi = 1 + w_\phi$, whereas the freezing models are characterized by the region between two curves: (c) $w'_\phi = 0.2w_\phi(1 + w_\phi)$ and (d) $w'_\phi = 3w_\phi(1 + w_\phi)$. The dotted line shows the border between the acceleration and deceleration of the field ($\ddot{\phi} = 0$), which corresponds to $w'_\phi = 3(1 + w_\phi)^2$.

where the upper and lower bounds correspond to the curves (c) and (d) in Fig. 7.9, respectively.

Let us next consider the allowed region of the thawing models. Taking the derivative of Eq. (7.6) with respect to the number of e-foldings N , we find [238]

$$w'_\phi = (1 + w_\phi)(3 + 2\xi + 3w_\phi), \quad (7.106)$$

where $\xi \equiv \ddot{\phi}/(H\dot{\phi})$. The border separating the acceleration and the deceleration of the scalar field corresponds to $\xi = 0$, i.e. the curve $w'_\phi = 3(1 + w_\phi)^2$ in the (w_ϕ, w'_ϕ) plane. In thawing models, the field is almost frozen during the matter era by the Hubble damping so that $w_\phi \sim -1$ and $\xi \ll 1$. This gives the following relation

$$w'_\phi \simeq 2\xi(1 + w_\phi). \quad (7.107)$$

The upper limit for ξ can be derived by using the condition $\ddot{\phi} \lesssim \dot{\phi}/t \approx (3/2)H\dot{\phi}$, where we have used the relation $H \approx 2/(3t)$ during the matter-dominated epoch.

This then provides the upper bound $\xi \lesssim 3/2$. Caldwell and Linder obtained the lower bound $\xi \gtrsim 1/2$ by numerically studying several classes of thawing models such as the PNGB model [182]. Hence the thawing models are characterized by the following region

$$1 + w_\phi \lesssim w'_\phi \lesssim 3(1 + w_\phi), \quad (7.108)$$

where the upper and lower bounds correspond to the curves (a) and (b) in Fig. 7.9, respectively.

In both freezing and thawing models the equation of state ranges in the region $-1 < w_\phi \lesssim -0.8$. While the observational data up to now are not sufficient to distinguish freezing and thawing models by the variation of w_ϕ , we may be able to do so with high-precision observations in the next decade.

7.6 Problems

- 7.1** Derive Eq. (7.25) for the linear perturbations about the fixed points for quintessence with an exponential (7.23) and then find the eigenvalues as given in the text.
- 7.2** Derive Eqs. (7.30) and (7.31) from the definition of x given in Eq. (7.29).
- 7.3** For the two-field quintessence with the potential $V(\phi_1, \phi_2) = V_1 e^{-\kappa \lambda_1 \phi_1} + V_2 e^{-\kappa \lambda_2 \phi_2}$, show that the effective equation of state w_{eff} for the stable fixed point responsible for cosmic acceleration is given by $w_{\text{eff}} = -1 + \lambda_{\text{eff}}^2/3$. Here λ_{eff} is defined by Eq. (7.48), i.e. $1/\lambda_{\text{eff}}^2 \equiv 1/\lambda_1^2 + 1/\lambda_2^2$.
- 7.4** Find the field potential (7.65) from the Kähler potential $K = [\ln(\kappa\varphi + \kappa\varphi^*)]^2/\kappa^2$ and the superpotential $W = M^{4+\alpha}\varphi^{-\alpha}$.
- 7.5** Derive the reconstruction equations (7.85) and (7.86) from Eqs. (7.7) and (7.8).

8

Dark energy as a modified form of matter II

In the previous section we have discussed quintessence as one of the basic examples of “modified matter models” of dark energy. There are other classes of modified matter models such as k-essence, phantoms, coupled dark energy, and unified models of dark energy and dark matter. These models are more complicated but also have phenomenologically rich properties. In this section we shall discuss these dark energy candidates.

8.1 k-essence

8.1.1 *k-essence models*

Quintessence is based on scalar field models using a canonical field with a slowly varying potential. It is known however that scalar fields with non-canonical kinetic terms often appear in the context of particle physics. The action for such models is in general given by

$$S = \int d^4x \sqrt{-g} \left[\frac{1}{2\kappa^2} R + P(\phi, X) \right] + S_m, \quad (8.1)$$

where $P(\phi, X)$ is a function in terms of a scalar field ϕ and its kinetic energy $X = -(1/2)g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi \equiv -(1/2)(\nabla\phi)^2$. The possibility to realize an accelerated expansion of the Universe according to the action (8.1) was suggested in Ref. [239] in the context of inflation. The application of this scenario to dark energy was first carried out in Ref. [42]. Later this analysis was extended to more general cases and the models based on the action (8.1) were named “k-essence” [43, 44]. The central point is that the cosmic acceleration can be realized by the *kinetic* energy X of the field ϕ .

The following models belong to k-essence.

- **(i) Low-energy effective string theory with derivative terms higher-order than X**

The low-energy effective string theory gives rise to higher-order derivative terms coming from α' and loop corrections to the tree-level action [64]. Let us consider the following low-energy effective string action in the presence of a derivative term $(\tilde{\nabla}\phi)^4$:

$$S = \frac{1}{2\kappa^2} \int d^4\tilde{x} \sqrt{-\tilde{g}} \left[B_g(\phi) \tilde{R} + B_\phi^{(0)}(\phi) (\tilde{\nabla}\phi)^2 + \alpha' c_1 B_\phi^{(1)}(\phi) (\tilde{\nabla}\phi)^4 + \mathcal{O}(\alpha'^2) \right], \quad (8.2)$$

where we set $V(\phi) = 0$ and $S_m = 0$. Note that we have used a tilde for quantities in the string frame unlike the action (7.75). Performing a conformal transformation, $g_{\mu\nu} = B_g \tilde{g}_{\mu\nu}$, the action (8.2) is transformed to the Einstein frame action (see Section 7.4.4):

$$S_E = \int d^4x \sqrt{-g} \left[\frac{1}{2\kappa^2} R + K(\phi) X + L(\phi) X^2 + \dots \right], \quad (8.3)$$

where

$$K(\phi) = 3 \left(\frac{1}{B_g} \frac{dB_g}{d\phi} \right)^2 - 2 \frac{B_\phi^{(0)}}{B_g}, \quad L(\phi) = 2c_1 \frac{\alpha'}{\kappa^2} B_\phi^{(1)}(\phi). \quad (8.4)$$

The action (8.3) belongs to the k -essence action (8.1) with

$$P = K(\phi) X + L(\phi) X^2. \quad (8.5)$$

- **(ii) Ghost condensate model**

In Section 7 we have seen that the current observations allow an equation of state w_{DE} smaller than -1 . It is possible to explain such an equation of state by considering a negative kinetic energy $-X$ with a field potential $V(\phi)$, which is called a phantom or a ghost field. However the phantom field is plagued by severe ultra-violet quantum instabilities because its energy density is not bounded from below. If we take into account the derivative term X^2 , the vacuum can be stabilized against the catastrophic particle production of ghosts and normal fields. This model, dubbed the ghost condensate model [240], has the Lagrangian density given by

$$P = -X + X^2/M^4, \quad (8.6)$$

where M is a constant having a dimension of mass. The model can be regarded as the specific case of (8.5), i.e. $K = -1$ and $L = 1/M^4$. As we will see later, there exists a de Sitter solution at $X/M^4 = 1/2$. There is also a modified version of the above model:

$$P = -X + e^{\kappa\lambda\phi} X^2/M^4, \quad (8.7)$$

which is called dilatonic ghost condensate model [241]. The correction of the type $e^{\kappa\lambda\phi} X^2/M^4$ can arise as a dilatonic higher-order correction to the tree-level string action, as we have discussed in case (i).

- **(iii) Tachyon field**

In string theories there exist unstable D-branes called non-Bogomol'nyi–Prasad–Sommerfield (BPS) D-branes besides stable BPS D-branes. These unstable branes are characterized by having a single tachyonic mode of a negative mass living on their world-volume. If we consider the dynamics of the tachyon on a non-BPS D3-brane, the effective 4-dimensional action is given by [242, 243]

$$P = -V(\phi)\sqrt{-\det(g_{\mu\nu} + \partial_\mu\phi\partial_\nu\phi)}, \quad (8.8)$$

where $V(\phi)$ is a field potential. In open string theory the tachyon starts to roll down from the top of the potential located at $\phi = 0$ and evolves toward a ground state at $|\phi| \rightarrow \infty$. The typical potential is $V(\phi) = V_0/\cosh(\beta\phi/2)$, where $\beta = 1$ for the bosonic string and $\beta = \sqrt{2}$ for the non-BPS D-brane in the superstring (in the unit of $\alpha' = 1$) [244]. This potential is too steep in the large $|\phi|$ region to sustain inflationary expansion of the Universe at late times. Tachyon potentials shallower than the inverse squared potential ($V(\phi) \propto \phi^{-2}$) can account for the late-time cosmic acceleration [245, 246, 247, 248]. There is another tachyon scenario based on a massive scalar field with mass m on the anti D3-brane, in which case the potential is given by $V(\phi) = V_0 e^{\frac{1}{2}m^2\phi^2}$ [249]. Since the potential energy is present at $\phi = 0$, it is possible to have the late-time cosmic acceleration in this scenario.

- **(iv) Dirac–Born–Infeld (DBI) theories**

The motion of a scalar field can be slowed down through the so-called “D-celeration” mechanism [250, 251] in which the field ϕ parametrizes a direction on the approximate Coulomb branch of the system in $\mathcal{N} = 4$ supersymmetric Yang–Mills theory. The speed of the field is restricted by the causality of the gravity side of the Anti de Sitter/Conformal Field Theory (AdS–CFT) correspondence [252]. This dynamics is well described by the DBI action for a probe D3-brane domain wall moving in the radial direction of the AdS₅ spacetime. The Lagrangian density describing this theory is given by [250, 251]

$$P = -f(\phi)^{-1}\sqrt{1 - 2f(\phi)\dot{X}} + f(\phi)^{-1} - V(\phi), \quad (8.9)$$

where $V(\phi)$ is a field potential and $f(\phi)$ is a warp factor of the AdS-like throat. For the AdS throat we have $f(\phi) = \lambda/\phi^4$, where λ is the 't Hooft coupling related to the Yang–Mills coupling g_{YM}^2 via the relation $\lambda = g_{\text{YM}}^2 N$ in the large N limit of the field theory. In the non-relativistic limit, $2f(\phi)\dot{X} \ll 1$, the Lagrangian density (8.9) reduces to $P = X - V(\phi)$, i.e. the one corresponding to a canonical scalar field. In the cosmological set up, it is possible to realize the accelerated expansion of the Universe even when the γ factor defined by $\gamma = 1/\sqrt{1 - f(\phi)\dot{\phi}^2}$ is much larger than unity (i.e. $f(\phi)\dot{\phi}^2 \simeq 1$). This situation is different from the tachyon field where the condition $\dot{\phi}^2 \ll 1$ is required for the cosmic acceleration. The application of the DBI theories to dark energy has been discussed in Refs. [253, 254].

8.1.2 Equation of state for *k-essence*

The energy-momentum tensor of the scalar field following from the action $S_\phi = \int d^4x \sqrt{-g} P(\phi, X)$ is given by

$$T_{\mu\nu}^{(\phi)} = -\frac{2}{\sqrt{-g}} \frac{\delta(\sqrt{-g}P)}{\delta g^{\mu\nu}} = P_{,X} \partial_\mu \phi \partial_\nu \phi + g_{\mu\nu} P, \quad (8.10)$$

where a suffix “ X ” or “ ϕ ” represents a partial derivative with respect to X or ϕ , respectively. The energy-momentum tensor of *k-essence* is that of a perfect fluid, $T_{\mu\nu} = (\rho + P)u_\mu u_\nu + g_{\mu\nu} P$, with velocity $\mu_\mu = \partial_\mu \phi / \sqrt{2X}$, pressure $P_\phi = P$, and energy density

$$\rho_\phi = 2X P_{,X} - P. \quad (8.11)$$

Then the equation of state of *k-essence* is

$$w_\phi = \frac{P_\phi}{\rho_\phi} = \frac{P}{2X P_{,X} - P}. \quad (8.12)$$

As long as the condition $|2X P_{,X}| \ll |P|$ is satisfied, w_ϕ can be close to -1 . In the flat FLRW background in the presence of a matter fluid with energy density ρ_M and pressure P_M , the Einstein equations are

$$3H^2 = \kappa^2 (\rho_\phi + \rho_M), \quad (8.13)$$

$$2\dot{H} = -\kappa^2 (2X P_{,X} + \rho_M + P_M), \quad (8.14)$$

$$\dot{\rho}_\phi + 3H(\rho_\phi + P_\phi) = 0. \quad (8.15)$$

For the ghost condensate model (8.6) we have

$$w_\phi = \frac{1 - X/M^4}{1 - 3X/M^4}, \quad (8.16)$$

which gives $-1 < w_\phi < -1/3$ for $1/2 < X/M^4 < 2/3$. In particular the de Sitter solution ($w_\phi = -1$) is realized at $X/M^4 = 1/2$. Since the field energy density is $\rho_\phi = M^4/4$ at the de Sitter point, it is possible to explain the present cosmic acceleration for $M \sim 10^{-3}$ eV.

In the flat FLRW spacetime the pressure and the energy density of the tachyon field follow from the action (8.8):

$$P_\phi = -V(\phi)\sqrt{1 - \dot{\phi}^2}, \quad \rho_\phi = \frac{V(\phi)}{\sqrt{1 - \dot{\phi}^2}}, \quad (8.17)$$

which give the equation of state

$$w_\phi = -1 + \dot{\phi}^2. \quad (8.18)$$

In order to realize $w_\phi \approx -1$ we require that $\dot{\phi}^2 \ll 1$. In the sense that the kinetic energy of the tachyon needs to be suppressed to realize cosmic acceleration, this scenario is different from k-essence. We simply classified the tachyon field as k-essence because it belongs to a class of the action (8.1).

Similarly the pressure and the energy density of the DBI field follow from the action (8.9):

$$P_\phi = \frac{\gamma - 1}{f\gamma} - V(\phi), \quad \rho_\phi = \frac{\gamma - 1}{f} + V(\phi), \quad (8.19)$$

where

$$\gamma = \frac{1}{\sqrt{1 - f(\phi)\dot{\phi}^2}}. \quad (8.20)$$

Hence the equation of state is

$$w_\phi = \frac{(\gamma - 1)/(f\gamma) - V(\phi)}{(\gamma - 1)/f + V(\phi)}. \quad (8.21)$$

In the slow-roll limit, $f(\phi)\dot{\phi}^2 \ll 1$, one has $w_\phi \rightarrow -1$. If we introduce the following variables

$$x_1 \equiv \frac{\kappa}{\sqrt{3}H} \sqrt{\frac{\gamma}{f}}, \quad x_2 \equiv \frac{\kappa\dot{\phi}\sqrt{\gamma}}{H}, \quad x_3 \equiv \frac{\kappa\sqrt{V}}{\sqrt{3}H}, \quad (8.22)$$

the equation of state (8.21) can be written as

$$w_\phi = \frac{[(1 - 1/\gamma)/\gamma]x_1^2 - x_3^2}{(1 - 1/\gamma)x_1^2 + x_3^2}, \quad (8.23)$$

where

$$\gamma = \frac{1}{\sqrt{1 - x_2^2/3x_1^2}}. \quad (8.24)$$

From Eq. (8.13) we obtain the constraint equation, $\Omega_M = 1 - (1 - 1/\gamma)x_1^2 - x_2^2$. In the case of the quadratic potential, $V(\phi) = (1/2)m_\phi^2\phi^2$, there is a fixed point that can be responsible for the cosmic acceleration [254]:

$$x_1 = \left[\mu(\sqrt{\mu^2 + 12} - \mu)/6 \right]^{1/2}, \quad x_2 = -\sqrt{3}x_1, \quad x_3 = \sqrt{3}(\sqrt{\mu^2 + 12} - \mu)/6,$$

$$w_{\text{eff}} = -1 + \mu(\sqrt{\mu^2 + 12} - \mu)/6, \quad \Omega_M = 0, \quad (8.25)$$

where w_{eff} is defined by $w_{\text{eff}} \equiv -1 - 2\dot{H}/(3H^2)$ and

$$\mu \equiv \frac{V_{,\phi}}{\kappa f^{1/2} V^{3/2}} = \frac{2\sqrt{2}}{\kappa\sqrt{\lambda}m_\phi}. \quad (8.26)$$

The cosmic acceleration occurs for $w_{\text{eff}} < -1/3$, i.e. $\mu < 2$. Since $x_2 = -\sqrt{3}x_1$, Eq. (8.24) shows that this fixed point exists in the ultra-relativistic region, $\gamma \rightarrow \infty$. Thus the DBI model allows inflationary solutions even if the field is not in the slow-roll regime.

In order to understand the cases in which the late-time cosmic acceleration is realized by the tracking of the *k-essence* field, it is important to derive conditions for the existence of scaling solutions where the energy density Ω_{DE} of dark energy relative to the energy density Ω_M of a background fluid remains constant (but non-zero). We will prove in Section 8.4.4 that the existence of scaling solutions restricts the form of the Lagrangian density to be

$$P(\phi, X) = Xg(Xe^{\kappa\lambda\phi}), \quad (8.27)$$

where λ is a constant and g is an arbitrary function in terms of $Y \equiv Xe^{\kappa\lambda\phi}$.

If we choose $g(Y) = 1 - c/Y$ then we obtain $p = X - ce^{-\kappa\lambda\phi}$, which corresponds to quintessence with an exponential potential. We have shown in Section 7.2 that this model in fact has a scaling fixed point (d) with $\Omega_\phi = 3(1 + w_M)/\lambda^2$. The dilatonic ghost condensate model belongs to the class of the Lagrangian density (8.27) by setting $g(Y) = -1 + Y/M^4$. One can also show that the tachyon field has a scaling solution for the potential $V(\phi) = M^6\phi^{-2}$ (see problem 8.1). The tachyon potential shallower than the inverse power-law potential enters a tracking regime [247, 248]. The DBI model also possesses scaling solutions for the potential $V(\phi) = (1/2)m_\phi^2\phi^2$ in the case of the AdS throat ($f(\phi) = \lambda/\phi^4$) [254].

8.1.3 Stability conditions for *k-essence*

In *k-essence* it can happen that the linear kinetic energy in X has a negative sign. Such a field, called a phantom or ghost scalar field [83], suffers from a quantum instability problem unless higher-order terms in X or ϕ are taken into account in the Lagrangian density. In the (dilatonic) ghost condensate scenario it is possible to avoid this quantum instability by the presence of the term X^2 .

Let us derive stability conditions of *k-essence* by considering small fluctuations $\delta\phi(t, \mathbf{x})$ around a background value $\phi_0(t)$ which is the solution in the FLRW spacetime. Then the field $\phi(t, \mathbf{x})$ can be decomposed in the form

$$\phi(t, \mathbf{x}) = \phi_0(t) + \delta\phi(t, \mathbf{x}). \quad (8.28)$$

Since we are interested in ultra-violet (UV) instabilities, it is not restrictive to consider a Minkowski background. Expanding $P(\phi, X)$ at the second order in $\delta\phi$ it is straightforward to find the Lagrangian and then the Hamiltonian for the

fluctuations. The perturbed Hamiltonian reads

$$\delta\mathcal{H} = (P_{,X} + 2XP_{,XX}) \frac{(\delta\phi)^2}{2} + P_{,X} \frac{(\nabla\delta\phi)^2}{2} - P_{,\phi\phi} \frac{(\delta\phi)^2}{2}. \quad (8.29)$$

The positive definiteness of the Hamiltonian is guaranteed if the following conditions hold

$$\xi_1 \equiv P_{,X} + 2XP_{,XX} \geq 0, \quad \xi_2 \equiv P_{,X} \geq 0, \quad (8.30)$$

$$\xi_3 \equiv -P_{,\phi\phi} \geq 0. \quad (8.31)$$

When we discuss the stability of classical perturbations, the quantity often used is the speed of sound c_s defined by [255]

$$c_s^2 \equiv \frac{P_{\phi,X}}{\rho_{\phi,X}} = \frac{\xi_2}{\xi_1}, \quad (8.32)$$

where we have used $P_\phi = P$ and $\rho_\phi = 2XP_{,X} - P$. In cosmological perturbation theory c_s^2 appears as a coefficient of the k^2/a^2 term, where a is the scale factor and k is the comoving wavenumber [90, 92]. While the classical fluctuations may be regarded as stable when c_s^2 is positive, the stability of quantum fluctuations requires both the conditions $\xi_1 \geq 0$ and $\xi_2 \geq 0$. These two conditions prevent an instability related to the presence of negative energy ghost states. If these conditions are violated, the vacuum is unstable under a catastrophic production of ghosts and photons pairs [256, 257]. The production rate from the vacuum is proportional to the phase space integral on all possible final states. Since only a UV cut-off can prevent the creation of modes of arbitrarily high energies, this is essentially a UV instability. The phantom model with the Lagrangian density $P(\phi, X) = -X - V(\phi)$ violates both the conditions $\xi_1 \geq 0$ and $\xi_2 \geq 0$, which means that the vacuum is unstable.

If we take into account higher-order terms such as X^2 in $P(\phi, X)$, it is possible to avoid the quantum instability mentioned above. Let us consider the dilatonic ghost condensate model with $P = -X + e^{\kappa\lambda\phi} X^2/M^4$. Since $\xi_1 = -1 + 6e^{\kappa\lambda\phi} X/M^4$ and $\xi_2 = -1 + 2e^{\kappa\lambda\phi} X/M^4$ in this case, the quantum instability is ensured for $e^{\kappa\lambda\phi} X/M^4 \geq 1/2$. The equation of state for the dilatonic ghost condensate is

$$w_\phi = \frac{1 - e^{\kappa\lambda\phi} X/M^4}{1 - 3e^{\kappa\lambda\phi} X/M^4}. \quad (8.33)$$

Hence we have $w_\phi \geq -1$ under the condition $e^{\kappa\lambda\phi} X/M^4 \geq 1/2$, which means that the phantom equation of state ($w_\phi < -1$) is not realized if we ensure the quantum stability. The tachyon model corresponds to $\xi_1 = V(\phi)(1 - 2X)^{-3/2} > 0$ and $\xi_2 = V(\phi)(1 - 2X)^{-1/2} > 0$, so that the quantum stability is ensured. The situation is similar for the DBI model in which $\xi_1 = (1 - 2f(\phi)X)^{-3/2} > 0$ and $\xi_2 = (1 - 2f(\phi)X)^{-1/2} > 0$.

The instability prevented by the condition (8.31) is of the tachyonic type and generally much less dramatic, as long as the conditions (8.30) are satisfied. If $P_{,\phi\phi} > 0$ there are large-scale modes with $k^2 < P_{,\phi\phi}/P_{,X}$, which undergo a classical exponential growth. This is an instability of the infra-red (IR) type as we often encounter cosmological perturbations of super-horizon modes in inflationary cosmology [94]. The ordinary cosmological (light) scalar field with a Lagrangian density $P = X - V(\phi)$ corresponds to $\xi_3 = -P_{,\phi\phi} = V_{,\phi\phi} \simeq H^2$. Even if the field has a negative mass squared such that $\xi_3 = V_{,\phi\phi} < 0$, the modes deep inside the Hubble radius ($k^2/a^2 \gg H^2$) are not subject to this negative instability. For the above reasons we shall adopt (8.30) but not (8.31) as fundamental criteria for the consistency of the theory on the physical scales we are interested in.

The sound speed of the field becomes superluminal ($c_s^2 > 1$) depending on the models of *k-essence*. This superluminal propagation is not favorable as the causality could be violated (although see Ref. [258]). Under the stability conditions (8.30) we find that the sound speed of *k-essence* does not exceed the speed of light for $2XP_{,XX} \geq 0$, i.e.

$$P_{,XX} \geq 0. \quad (8.34)$$

Since $P_{,XX} = V(\phi)(1 - 2X)^{-3/2} > 0$ for the tachyon field and $P_{,XX} = f(\phi)(1 - 2f(\phi)X)^{-3/2} > 0$ for the DBI field, the superluminal propagation does not occur in these models. This is also the case for the dilatonic ghost condensate model in which $P_{,XX} = 2e^{\kappa\lambda\phi}/M^4 > 0$. If we consider the model of the type $P = c_1X + c_2X^2$, where c_1 and c_2 are constants, we have that $P_{,XX} = 2c_2$. Hence the sound speed becomes superluminal for $c_2 < 0$. As we will see in Section 8.1.5, this situation arises for *k-essence* models that aim to solve the coincidence problem of dark energy.

8.1.4 Cosmological dynamics for the dilatonic ghost condensate

As an example of *k-essence*, let us study the cosmological dynamics of the dilatonic ghost condensate model (8.7) in the flat FLRW background. As a matter fluid we take into account both non-relativistic matter (energy density ρ_m) and radiation (energy density ρ_r). Since $P_\phi = -X + e^{\kappa\lambda\phi}X^2/M^4$ and $\rho_\phi = -X + 3e^{\kappa\lambda\phi}X^2/M^4$ in Eqs. (8.13), (8.14), and (8.15), we obtain

$$3H^2 = \kappa^2 \left(-\frac{1}{2}\dot{\phi}^2 + \frac{3}{4}e^{\kappa\lambda\phi} \frac{\dot{\phi}^4}{M^4} + \rho_m + \rho_r \right), \quad (8.35)$$

$$2\dot{H} = \kappa^2 \left(\dot{\phi}^2 - e^{\kappa\lambda\phi} \dot{\phi}^4/M^4 - \rho_m - 4\rho_r/3 \right), \quad (8.36)$$

$$\ddot{\phi}(3e^{\kappa\lambda\phi}\dot{\phi}^2/M^4 - 1) + 3H\dot{\phi}(e^{\kappa\lambda\phi}\dot{\phi}^2/M^4 - 1) + 3\kappa\lambda e^{\kappa\lambda\phi}\dot{\phi}^4/(4M^4) = 0. \quad (8.37)$$

Defining the following variables

$$x_1 \equiv \frac{\kappa \dot{\phi}}{\sqrt{6}H}, \quad x_2 \equiv \frac{\dot{\phi}^2 e^{\kappa\lambda\phi}}{2M^4}, \quad x_3 \equiv \frac{\kappa \sqrt{\rho_r}}{\sqrt{3}H}, \quad (8.38)$$

we find that these satisfy the following autonomous equations

$$\frac{dx_1}{dN} = -x_1 \frac{6(2x_2 - 1) + 3\sqrt{6}\lambda x_1 x_2}{2(6x_2 - 1)} + \frac{x_1}{2}(3 - 3x_1^2 + 3x_1^2 x_2 + x_3^2), \quad (8.39)$$

$$\frac{dx_2}{dN} = x_2 \frac{3x_2(4 - \sqrt{6}\lambda x_1) - \sqrt{6}(\sqrt{6} - \lambda x_1)}{1 - 6x_2}, \quad (8.40)$$

$$\frac{dx_3}{dN} = \frac{x_3}{2}(-1 - 3x_1^2 + 3x_1^2 x_2 + x_3^2), \quad (8.41)$$

together with

$$w_{\text{eff}} = -1 - \frac{2\dot{H}}{3H^2} = -x_1^2 + x_1^2 x_2 + \frac{1}{3}x_3^2, \quad w_\phi = \frac{P_\phi}{\rho_\phi} = \frac{1 - x_2}{1 - 3x_2},$$

$$\Omega_\phi = -x_1^2 + 3x_1^2 x_2, \quad \Omega_r = x_3^2, \quad \Omega_m = 1 + x_1^2 - 3x_1^2 x_2 - x_3^2. \quad (8.42)$$

In order to ensure the quantum stability we require that $x_2 \geq 1/2$. The following fixed points are relevant for viable cosmological evolution:

- (a) Radiation point: $(x_1, x_2, x_3) = (0, 1/2, 1)$.
This satisfies $w_{\text{eff}} = 1/3$, $w_\phi = -1$, $\Omega_\phi = 0$, $\Omega_r = 1$, and $\Omega_m = 0$.
- (b) Matter point: $(x_1, x_2, x_3) = (0, 1/2, 0)$.
This satisfies $w_{\text{eff}} = 0$, $w_\phi = -1$, $\Omega_\phi = 0$, $\Omega_r = 0$, and $\Omega_m = 1$.
- (c) Accelerated point: $(x_1, x_2, x_3) = (-\sqrt{6}\lambda f_-(\lambda)/4, 1/2 + \lambda^2 f_+(\lambda)/16, 0)$, where

$$f_\pm(\lambda) \equiv 1 \pm \sqrt{1 + 16/(3\lambda^2)}. \quad (8.43)$$

This satisfies $w_{\text{eff}} = w_\phi = (-8 + \lambda^2 f_+(\lambda))/(8 + 3\lambda^2 f_+(\lambda))$, $\Omega_\phi = 1$, $\Omega_r = 0$, and $\Omega_m = 0$. The cosmic acceleration occurs for $-1 \leq w_{\text{eff}} < -1/3$, which translates into the condition

$$0 \leq \lambda < \sqrt{6}/3. \quad (8.44)$$

One can also show that this accelerated point is stable for $0 \leq \lambda < \sqrt{3}$ [241]. Hence the stability of the accelerated point is ensured under the condition (8.44).

We also have other fixed points. For example there is another accelerated point $(x_1, x_2, x_3) = (-\sqrt{6}\lambda f_+(\lambda)/4, 1/2 + \lambda^2 f_-(\lambda)/16, 0)$, but this corresponds to the quantum instability region $x_2 < 1/2$ (i.e. the phantom equation of state $w_\phi < -1$). During the matter era we also have the scaling solution with $x_1 = \sqrt{6}/(2\lambda)$, $x_2 = 1$, $x_3 = 0$, $\Omega_\phi = 3/\lambda^2$, and $w_\phi = 0$. However the existence of a viable scaling matter

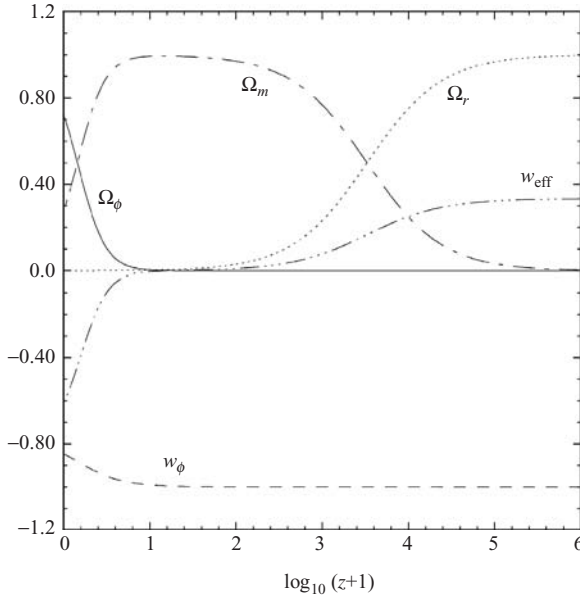


Figure 8.1 Evolution of Ω_ϕ , Ω_m , Ω_r , w_{eff} , and w_ϕ for the dilatonic ghost condensate model (8.7) with $\lambda = 0.2$ versus the redshift z . Initially the equation of state w_ϕ is close to -1 because x_2 is close to $1/2$. A deviation from $w_\phi = -1$ appears in the low-redshift regime. Initial conditions are chosen to be $x_1 = 6.0 \times 10^{-11}$, $x_2 - 1/2 = 1.0 \times 10^{-9}$, and $x_3 = 0.999$ at $\log_{10}(z + 1) = 6.218$.

era requires the condition $\lambda > \sqrt{3}$, which is not compatible with the condition (8.44).

In Fig. 8.1 we plot the cosmological evolution of the dilatonic ghost condensate model with $\lambda = 0.2$. The initial conditions at the radiation era are chosen to be close to the radiation point $(x_1, x_2, x_3) = (0, 1/2, 1)$ with $x_2 > 1/2$ (in order to ensure the stability of quantum fluctuations). In Fig. 8.1 we find that the successful cosmological evolution is in fact realized. Since $x_2 \approx 1/2$ during radiation and most matter eras, the equation of state w_ϕ is close to -1 . In these epochs the energy density of the field is given by $\rho_\phi \approx \dot{\phi}^2/4$. Since $x_1 = \kappa \dot{\phi}/(\sqrt{6}H) \approx 0$ at the radiation and matter fixed points, ρ_ϕ is negligibly small relative to the background fluid density. The field energy density begins to dominate over the background fluid density after x_1 grows to the order of unity. At this epoch the deviation from $x_2 = 1/2$ becomes important so that the solutions approach the accelerated point (c). When $\lambda = 0.2$ we have $w_\phi (= w_{\text{eff}}) = -0.788$ for the point (c). In the numerical simulation of Fig. 8.1 the equation of state today (corresponding to $\Omega_\phi \simeq 0.72$) is $w_\phi \simeq -0.847$. Since the deviation from $w_\phi = -1$ appears around the present epoch, the dilatonic ghost condensate model corresponds to the thawing model

of k-essence. Unlike thawing models of *quintessence* the field acquires a nearly constant energy by its kinetic term.

Finally we recall that the sound speed of the dilatonic ghost condensate model is smaller than the speed of light because the condition (8.34) holds. The sound speed squared in this case is given by

$$c_s^2 = \frac{2x_2 - 1}{6x_2 - 1}. \quad (8.45)$$

The condition (8.44) for the existence of the late-time accelerated point gives $1/2 \leq x_2 < 2/3$. Hence the sound speed is in the region

$$0 \leq c_s < 1/3, \quad (8.46)$$

which means that this model does not violate causality.

8.1.5 k-essence and the coincidence problem

We have shown that the dilatonic ghost condensate model gives rise to successful cosmological evolution while satisfying the conditions of quantum stability and the sound speed. Recall that in this model the energy density of the field during radiation and matter eras is negligibly small relative to the background fluid density. From the viewpoint of alleviating the coincidence problem of dark energy, a satisfactory k-essence model should allow a cosmological evolution in which the solutions finally approach the accelerating phase even if they start from relatively large values of the k-essence energy density Ω_ϕ at the radiation era. In the following we shall discuss such a possibility. Let us notice that, just as for the tracker solutions, the global future attractor for k-essence is $\Omega_\phi = 1$ and therefore it cannot answer the question of why only today we observe matter and dark energy with similar densities.

Let us consider the k-essence model with the Lagrangian density

$$P(\phi, X) = K(\phi)p(X). \quad (8.47)$$

For later convenience we define the following quantities

$$g(y) \equiv p(X)y, \quad y \equiv 1/\sqrt{X}. \quad (8.48)$$

In order to have a scaling solution with $w_\phi = w_r$ at the early epoch of the radiation era, Armendariz-Picon *et al.* [43, 44] have chosen the form $K(\phi) = 1/\phi^2$ (in the unit of $\kappa^2 = 1$). In fact this form can be obtained by using the scaling Lagrangian (8.27), see problem 6.1. Depending on the form of $g(Y)$, the k-essence scaling solution is not necessarily stable during the radiation era so that the initial scaling epoch is followed by other fixed points of the system. For the choice $K(\phi) = 1/\phi^2$ the k-essence energy density is given by $\rho_\phi = -g'(y)/\phi^2$, where $g'(y) \equiv dg/dy$.

Since $\rho_\phi > 0$, it follows that

$$g'(y) < 0. \quad (8.49)$$

Since the quantities ξ_1 and ξ_2 defined in Eq. (8.30) are given by $\xi_1 = y^3 g''/(2\phi^2)$ and $\xi_2 = Ky(g - yg')/\phi^2$, respectively, the conditions for the quantum stability require that

$$g''(y) > 0, \quad g(y) - yg'(y) > 0. \quad (8.50)$$

The condition for the avoidance of the superluminal propagation of the sound speed of *k-essence* corresponds to $P_{,XX} = -y^3(g - yg' - y^2 g'')/(4\phi^2) \geq 0$, which translates into

$$g(y) - yg'(y) \leq y^2 g''(y). \quad (8.51)$$

While we have restricted the case where $K(\phi) = 1/\phi^2$, the conditions (8.49), (8.50), and (8.51) persist for any positive function of $K(\phi)$.

The equation of state w_ϕ and the sound speed squared c_s^2 of *k-essence* are given by

$$w_\phi = -\frac{g(y)}{yg'(y)}, \quad c_s^2 = \frac{g(y) - yg'(y)}{y^2 g''(y)}. \quad (8.52)$$

The existence of the late-time accelerated solution requires the condition $w_\phi < 0$, which implies that $g(y) < 0$ today. In addition to *k-essence* we take into account the radiation and non-relativistic matter in Eqs. (8.13) and (8.14), i.e. $\rho_M = \rho_r + \rho_m$, $P_M = \rho_r/3$. Using Eq. (8.15) as well, we get the following equations for $K(\phi) = 1/\phi^2$ (see problem 8.2):

$$\frac{dy}{dN} = \frac{2\sqrt{-6g'(y)}}{yg''(y)} \left[r(y) - \sqrt{\Omega_\phi} \right], \quad (8.53)$$

$$\frac{d\Omega_\phi}{dN} = 3\Omega_\phi(1 - \Omega_\phi)(w_{rm} - w_\phi), \quad (8.54)$$

$$\frac{dw_{rm}}{dN} = w_{rm}(3w_{rm} - 1), \quad (8.55)$$

where

$$r(y) \equiv \frac{\sqrt{6}}{4} \frac{g(y) - yg'(y)}{\sqrt{-g'(y)}}, \quad \Omega_\phi \equiv \frac{\rho_\phi}{3H^2}, \quad w_{rm} \equiv \frac{\rho_r}{3(\rho_r + \rho_m)}. \quad (8.56)$$

From the above equations the fixed points of the system are given by $\Omega_\phi = r^2(y_f)$ with either $\Omega_\phi = 0$, or $\Omega_\phi = 1$, or $w_\phi(y_f) = w_{rm}$. In the radiation-dominated epoch ($\rho_r \gg \rho_m$) one has $w_{rm} = 1/3$, whereas after the radiation domination $w_{rm} = 0$. Both correspond to the fixed points of Eq. (8.55). We have the following fixed points relevant to *k-essence* models that aim to solve the coincidence problem.

- (i) Radiation era

The solutions start from the scaling radiation fixed point with $w_\phi = w_{rm} = 1/3$ and $\Omega_\phi(y_r) = r^2(y_r) \lesssim 0.045$ [see the BBN bound (7.50)].

- (ii) Matter era

During the matter era we have two possible fixed points. Which fixed points are reached depends on the form of $g(y)$ as well as initial conditions.

- (a) de Sitter fixed point

This satisfies $w_{rm} = 0$ and $\Omega_\phi = r^2(y_m) \simeq 0$. The condition $r \simeq 0$ gives $g \simeq yg'$ from Eq. (8.56) and hence $w_\phi = -1$ from Eq. (8.52).

- (b) Dust fixed point

This corresponds to the dust scaling solution with $w_\phi = w_{rm} = 0$ and $\Omega_\phi(y_m) = r^2(y_m) \neq 0$.

- (iii) Accelerated era

The solutions finally reach the k-essence accelerated attractor with $w_{rm} = 0$, $-1 < w_\phi < 0$, and $\Omega_\phi = r^2(y_k) \simeq 1$.

A few examples of k-essence models satisfying the conditions (8.49) and (8.50) are [43, 44]

$$P = \frac{1}{\phi^2} \left(-2.01 + 2\sqrt{1+X} + 3 \cdot 10^{-17} X^3 - 10^{-24} X^4 \right), \quad (8.57)$$

and

$$P = \frac{1}{\phi^2} \left(-2.05 + 2\sqrt{1+X - 10^{-8}X^2 + 10^{-12}X^3 - 10^{-16}X^4 + 10^{-20}X^5 - 10^{-24}(X/2)^6} \right). \quad (8.58)$$

In Fig. 8.2 the evolution of Ω_ϕ , w_ϕ , and c_s^2 is plotted for the model (8.57) as a function of the redshift z . The solution starts from the radiation scaling attractor with $\Omega_\phi \simeq 0.02$ and $w_\phi \simeq 1/3$. This is followed by a phase of decreasing Ω_ϕ associated with the growth of $w_\phi (> 1/3)$, see Eq. (8.54). The solution then approaches the de Sitter point with $w_\phi = -1$ and $\Omega_\phi = r^2(y_m) \simeq 0$ around the beginning of the matter-dominated epoch, which leads to the growth of Ω_ϕ . The solution finally approaches the k-essence attractor with $\Omega_\phi \simeq 1$ and $-1 < w_\phi < 0$. Since the present epoch is on the way from the de Sitter point to the k-essence attractor, the evolution of the k-essence equation of state w_ϕ is similar to that for the dilatonic ghost condensate model in the low-redshift region. Note that the model (8.58) gives rise to similar cosmological evolution.

Malquarti *et al.* [259] studied the basins of attraction of the radiation scaling solution for both models (8.57) and (8.58) and showed that they are restricted in

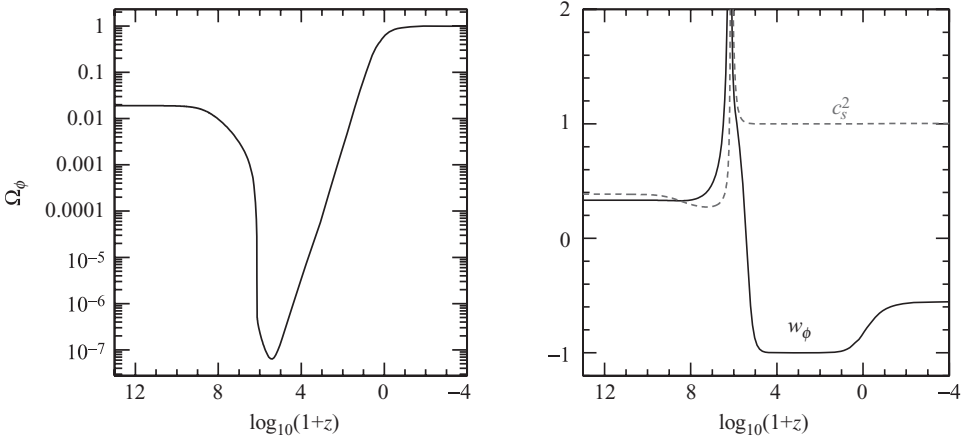


Figure 8.2 An example of the cosmological dynamics for the model (8.57) constructed to solve the coincidence problem of dark energy. The left panel shows the evolution of the k -essence energy Ω_ϕ , whereas the right panel plots the k -essence equation of state w_ϕ and the sound speed squared c_s^2 . While the solution finally approaches an accelerated point at late times even if Ω_ϕ is large initially, it is inevitable to avoid the superluminal propagation of the sound speed. From Ref. [260].

a very small region in the (y, Ω_ϕ) plane. For the model (8.57), for example, the solutions are not attracted by the radiation scaling solution for initial conditions of Ω_ϕ smaller than 0.01. Most of the solutions typically approach the k -essence attractor after an increase of the scale factor by about a factor 10^5 . This early dominance of the k -essence attractor does not give rise to the matter-dominated epoch and hence this case is excluded.

It is important to notice that the k -essence models constructed to solve the coincidence problem inevitably give rise to the superluminal propagation of the field ($c_s^2 > 1$) at some stage of the cosmological evolution [260]. The argument is as follows. The dynamics of the k -essence models to solve the coincidence problem starts from the radiation scaling solution with $w_\phi(y_r) = 1/3$, $\Omega_\phi = r^2(y_r) \lesssim 0.045$ and the solutions finally approach the k -essence attractor with $-1 < w_\phi(y_k) < 0$, $\Omega_\phi = r^2(y_k) \simeq 1 > r^2(y_r)$. Using the relation $w_\phi = -g(y)/yg'(y)$ together with the condition (8.49), it follows that $g(y_r) > 0$ and $g(y_k) < 0$ to realize $w_\phi(y_r) = 1/3$ and $w_\phi(y_k) < 0$, respectively. Since $g(y)$ is a monotonically decreasing function, we always have $y_k > y_r$. Remembering that $r(y_k) > r(y_r)$, r needs to grow when y increases from y_r to y_k .

Taking the derivative of $r(y)$ with respect to y , we find

$$r'(y) = \frac{\sqrt{6}}{8} \frac{yg''(y)}{\sqrt{-g'(y)}} [w_\phi(y) - 1]. \quad (8.59)$$

Using the condition $g''(y) > 0$, the increase of $r(y)$ occurs only for $w_\phi(y) > 1$. Hence there is an interval between y_r and y_k in which $w_\phi(y) > 1$. Since the equation of state of the k-essence attractor corresponds to $w_\phi(y_k) < 0$, there exists also an interval in which $w'_\phi(y) < 0$ and $w_\phi(y) > 1$. From Eq. (8.52) the derivative of w_ϕ with respect to y is given by

$$w'_\phi(y) = \frac{g(y)g'(y) + yg(y)g''(y) - yg'(y)^2}{y^2g'(y)^2}. \quad (8.60)$$

The conditions $w'_\phi(y) < 0$ and $w_\phi(y) > 1$ imply that $yg g'' < yg'^2 - gg'$ and $g > -yg'$, respectively. Since $yg'' > 0$ the second inequality gives $yg g'' > -y^2g'g''$. Substituting this into the first inequality, we obtain $g'(yg' - g + y^2g'') > 0$ and hence $g - yg' > y^2g''$. This corresponds to the violation of the condition (8.51). Thus we have shown that the k-essence models constructed to alleviate the coincidence problem lead to the superluminal propagation of the field sound speed.

In addition to the above-mentioned problems of initial conditions and the superluminal propagation of the sound speed, we should mention that the construction of models such as (8.57) and (8.58) is difficult in the framework of particle physics.

8.2 Phantoms

The current observations allow the possibility of the equation of state w_{DE} smaller than -1 , which is generally referred to as a *phantom* equation of state. As we see from Eq. (8.12), the k-essence model with a positive energy density ($\rho_\phi > 0$) gives rise to the equation of state $w_\phi < -1$ for

$$P_{,X} < 0. \quad (8.61)$$

The simplest model to satisfy this condition is a scalar field with a negative kinetic energy [83], i.e.

$$P(X, \phi) = -X - V(\phi), \quad (8.62)$$

where $V(\phi)$ is the field potential. This is called a phantom or ghost scalar field and its energy density and pressure are given by $\rho_\phi = -\dot{\phi}^2/2 + V(\phi)$ and $P_\phi = -\dot{\phi}^2/2 - V(\phi)$, respectively. The equation of state of the phantom field is

$$w_\phi = \frac{\dot{\phi}^2/2 + V(\phi)}{\dot{\phi}^2/2 - V(\phi)}. \quad (8.63)$$

One has $w_\phi < -1$ for $\dot{\phi}^2/2 < V(\phi)$.

The cosmological dynamics of the phantom scalar field has been discussed in a number of papers [256, 261, 262]. The phantom field rolls *up* the potential because of the negative kinetic energy. If the potential is unbounded from above, the field

energy density continues to grow toward infinity. In the case of an exponential potential $V(\phi) = V_0 e^{-\kappa\lambda\phi}$ the equation of state w_ϕ approaches a constant value $w_\phi = -1 - \lambda^2/3 < -1$, which results in a big-rip singularity. If the potential is bounded from above as in the bell types such as $V(\phi) = V_0 e^{-\phi^2/\sigma^2}$ [256] and $V(\phi) = V_0 [\cosh(\beta\phi/m_{\text{pl}})]^{-1}$ [261], the solutions evolve from the phantom equation of state ($w_\phi < -1$) and then approach $w_\phi = -1$ as the field settles at the potential *maximum*.

Although the above cosmological dynamics might be acceptable classically, the phantom Lagrangian density corresponds to $\xi_1 = P_{,X} + 2XP_{,XX} = -1$ and $\xi_2 = P_{,X} = -1$ in Eq. (8.30). This means that the stability conditions (8.30) are not satisfied, which leads to an unstable vacuum state [256, 257]. In order to avoid the catastrophic instability of the vacuum, we need to consider theories in which the interaction between ghosts and normal fields is as weak as possible. Even if decoupled from the matter fields, ghosts couple to gravitons which mediate the vacuum decay process of the form [257]:

$$\text{Vacuum} \rightarrow 2\phi + 2\gamma, \quad (8.64)$$

where ϕ is a ghost field. This corresponds to a spontaneous creation of a ghost pair and a photon pair. The divergence of the phase-space integral can be avoided only if we impose a Lorentz non-invariant momentum space cut-off at the scale Λ_c with the creation rate

$$\Gamma \sim \Lambda_c^8 / m_{\text{pl}}^4, \quad (8.65)$$

where $\Gamma = \Gamma_{0 \rightarrow 2\phi + 2\gamma}$.

In order to derive the density of photons which are generated by the vacuum decay (8.64), we need to solve the equation for the phase space number density n of ghosts and photons in an expanding Universe:

$$\frac{d}{dt}(a^3 n) = a^3 \Gamma. \quad (8.66)$$

When the evolution of the scale factor is given by $a \propto t^p$, we obtain the following solution

$$n(t) = \Gamma \frac{t}{3p+1} = \frac{\Gamma}{H} \frac{p}{3p+1}. \quad (8.67)$$

The number density in the present Universe is estimated to be $n(t_0) \sim \Gamma/H_0 \sim \Lambda_c^8/(m_{\text{pl}}^4 H_0)$. Since the energy spectrum is peaked around $E \sim \Lambda_c$, we find

$$\frac{dn}{dE} \sim \frac{\Lambda_c^7}{m_{\text{pl}}^4 H_0}. \quad (8.68)$$

The spectrum of photons with energy E around Λ_c has been constrained by observations of the diffuse gamma ray background. The EGRET experiment measured the derivative of the photon flux F with respect to E to be [263]

$$\frac{dF}{dE} = (7.32 \pm 0.34) \times 10^{-9} \left(\frac{E}{451 \text{ MeV}} \right)^{-2.1} (\text{cm}^2 \text{ s sr MeV})^{-1}. \quad (8.69)$$

From the requirement that Eq. (8.68) does not exceed the limit given in Eq. (8.69), we obtain the following upper bound

$$\Lambda_c \lesssim 3 \text{ MeV}. \quad (8.70)$$

This means that the ghost field must originate from new physics far below the TeV scale. Hence the standard high-energy field theory is not likely to be responsible for the low-energy effective ghosts. Instead we are forced to imagine that the ghost comes from a low-energy sector completely hidden from the standard model, except for the gravitational interaction. So far no consistent theories for such effective ghosts have been constructed in the framework of particle physics. We also stress that the bound (8.70) has been derived under the minimal coupling of gravity. The presence of other couplings can even lead to tighter bounds.

We also note that even at the classical level the presence of the ghost field is problematic. One would expect in fact that the Universe becomes very strongly anisotropic with the anisotropic energy density (the positive energy density of large-scale gravitational waves) being compensated by a negative energy density of the ghost field. Therefore it does not successfully explain the observed isotropy and homogeneity of the Universe [264].

In Fig. 7.7 we find that the current observations allow the evolution of the dark energy equation of state crossing $w_{\text{DE}} = -1$. A model that might cross the cosmological constant boundary is dubbed *quintom* in Ref. [265]. It was shown in Ref. [266] that such a crossing is possible with a multi-field model (phantom + normal fields), i.e. the Lagrangian density $P = (1/2)(\nabla\phi_1)^2 - V_1(\phi_1) - (1/2)(\nabla\phi_2)^2 - V_2(\phi_2)$. However this model also suffers from the instability problem mentioned above because of the negative kinetic energy. The phantom equation of state ($w_{\text{DE}} < -1$) as well as the cosmological boundary crossing can be realized in $f(R)$ gravity [267, 368, 371], scalar tensor theory [268, 269, 264], and Lorentz violating models [270] without introducing a scalar field with a negative kinetic energy, while avoiding the quantum instability. In this sense it is not necessary to introduce a phantom field to produce a dark energy equation of state w_{DE} smaller than -1 .

8.3 Coupled dark energy

The fact that the energy density of dark energy is the same order as that of dark matter in the present Universe suggests that there may be some relation between them. We discuss such coupled dark energy scenarios in this subsection.

Several different forms of the coupling between dark energy and dark matter have been proposed. One possibility is to consider an interaction between a quintessence field ϕ and dark matter with an interaction of the form $Q\rho_m\dot{\phi}$ [16, 17]. In fact this type of interaction appears in the context of scalar-tensor theories (including Brans–Dicke theory) [56, 40, 271], $f(R)$ gravity [272], and dilaton gravity [211] after a conformal transformation to the Einstein frame. In Brans–Dicke theory, for example, a coupling between a scalar field ϕ and a Ricci scalar R gives rise to a constant coupling Q between ϕ and a non-relativistic matter in the Einstein frame [273]. Another approach is to introduce an interaction of the form $\Gamma\rho_m$ on the r.h.s. of the continuity equations (ρ_m is the dark matter energy density) with the normalization of Γ in terms of the Hubble parameter H , i.e. $\Gamma/H = \delta$, where δ is a dimensionless coupling [274, 275, 276, 277, 278, 279, 280]. This is basically a fluid description of coupled dark energy.

In the following we start from the coupled quintessence scenario and then proceed to coupled dark energy models with an interaction of the form $\delta H\rho_m$, where δ is a dimensionless coupling. We also discuss the mass varying neutrino scenario and the coupling of a scalar field with an electromagnetic field.

8.3.1 Coupled quintessence with an exponential potential

Let us consider an interaction between a scalar field ϕ and a non-relativistic matter in the form

$$\nabla_\mu T_{\nu(\phi)}^\mu = -Q T_M \nabla_\nu \phi, \quad \nabla_\mu T_{\nu(M)}^\mu = +Q T_M \nabla_\nu \phi, \quad (8.71)$$

where $T_{\nu(\phi)}^\mu$ and $T_{\nu(M)}^\mu$ are the energy-momentum tensors of ϕ and non-relativistic matter, respectively, with a trace $T_M = -\rho_M + 3P_M$ of the matter fluid. Since the radiation is traceless ($\rho_M = 3P_M$), the coupling-dependent terms vanish in Eq. (8.71). Meanwhile non-relativistic matter such as dark matter and baryons have direct couplings with the scalar field ϕ .

Generally the coupling strength Q of baryons is different from that of dark matter [17, 281, 282, 271]. If we assume the baryons to be completely uncoupled they follow geodesics (i.e. they are free of any long-range force beside gravity) and we can directly compare the results with observations, since generally speaking in any (classical) experiment we assume our equipment (rods, clocks, etc) not to possess

long-range interactions beside gravity. We say that the frame in which baryons follow geodesics is the “physical” frame, meaning simply that we can directly compare results with observations. If on the contrary the baryons are coupled then the physical frame has to be obtained through a conformal transformation, as we will discuss in Section 8.4.

Although we assume the baryons to be uncoupled, as far as cosmology is concerned this makes generally only a small difference, since baryons are subdominant. Therefore for simplicity we discuss here a single matter fluid with an universal coupling. We discuss in the next section a case in which uncoupled baryons can lead to a considerable difference, e.g., the presence of a baryon-dominated epoch. We also assume that the coupling Q is constant. A constant coupling Q arises in Brans–Dicke theory after a conformal transformation to the Einstein frame, as we will see in Section 9.2. In this section we shall use the unit $\kappa^2 = 1$ unless otherwise stated.

The field Lagrangian density of the coupled quintessence is $\mathcal{L}_\phi = -(1/2)g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi) + \mathcal{L}_{\text{int}}$, where the part $\mathcal{L}_{\text{int}}$ gives rise to the interacting energy-momentum tensor given in Eq. (8.71).¹ For the field potential $V(\phi)$ we can take the exponential type

$$V(\phi) = V_0 e^{-\lambda\phi}, \quad (8.72)$$

although of course other choices can be made [271]. Without losing generality the constant λ can be assumed to be positive. For the interaction given in Eq. (8.71), the field ϕ , non-relativistic matter, and radiation obey the following equations of motion, respectively, in the flat FLRW background:

$$\dot{\rho}_\phi + 3H(\rho_\phi + P_\phi) = -Q\rho_m\dot{\phi}, \quad (8.73)$$

$$\dot{\rho}_m + 3H\rho_m = +Q\rho_m\dot{\phi}, \quad (8.74)$$

$$\dot{\rho}_r + 4H\rho_r = 0, \quad (8.75)$$

together with the usual Friedmann equation

$$3H^2 = \rho_\phi + \rho_m + \rho_r. \quad (8.76)$$

Since $\rho_\phi = (1/2)\dot{\phi}^2 + V(\phi)$ and $P_\phi = (1/2)\dot{\phi}^2 - V(\phi)$, Eq. (8.73) can be written as

$$\ddot{\phi} + 3H\dot{\phi} + V_{,\phi} = -Q\rho_m. \quad (8.77)$$

In order to study the dynamics of the system we introduce the following variables

$$x_1 \equiv \frac{\dot{\phi}}{\sqrt{6}H}, \quad x_2 \equiv \frac{\sqrt{V}}{\sqrt{3}H}, \quad x_3 \equiv \frac{\sqrt{\rho_r}}{\sqrt{3}H}. \quad (8.78)$$

¹ Although we focus on the coupled quintessence, it is possible to consider a coupling between dark matter and a k-essence field. See Ref. [283] for cosmological dynamics of coupled k-essence fields.

Table 8.1 *The fixed points for the coupled quintessence model with an exponential potential.*

Name	x_1	x_2	x_3	Ω_ϕ	Ω_r	w_ϕ	w_{eff}
(a)	$-\frac{\sqrt{6}Q}{3}$	0	0	$\frac{2Q^2}{3}$	0	1	$\frac{2Q^2}{3}$
(b1)	1	0	0	1	0	1	1
(b2)	-1	0	0	1	0	1	1
(c)	$\frac{\lambda}{\sqrt{6}}$	$(1 - \frac{\lambda^2}{6})^{1/2}$	0	1	0	$-1 + \frac{\lambda^2}{3}$	$-1 + \frac{\lambda^2}{3}$
(d)	$\frac{\sqrt{6}}{2(Q+\lambda)}$	$[\frac{2Q(Q+\lambda)+3}{2(Q+\lambda)^2}]^{1/2}$	0	$\frac{Q(Q+\lambda)+3}{(Q+\lambda)^2}$	0	$\frac{-Q(Q+\lambda)}{Q(Q+\lambda)+3}$	$\frac{-Q}{Q+\lambda}$
(e)	0	0	1	0	1	—	$\frac{1}{3}$
(f)	$-\frac{1}{\sqrt{6}Q}$	0	$(1 - \frac{1}{2Q^2})^{1/2}$	$\frac{1}{6Q^2}$	$1 - \frac{1}{2Q^2}$	1	$\frac{1}{3}$
(g)	$\frac{2\sqrt{6}}{3\lambda}$	$\frac{2\sqrt{3}}{3\lambda}$	$(1 - \frac{4}{\lambda^2})^{1/2}$	$\frac{4}{\lambda^2}$	$1 - \frac{4}{\lambda^2}$	$\frac{1}{3}$	$\frac{1}{3}$

Taking the derivative of Eq. (8.76) in terms of the number of e-foldings N together with the use of Eqs. (8.73)–(8.75), we obtain

$$\frac{1}{H} \frac{dH}{dN} = -\frac{1}{2} (3 + 3x_1^2 - 3x_2^2 + x_3^2). \quad (8.79)$$

The effective equation of state defined in Eq. (4.96) is

$$w_{\text{eff}} = x_1^2 - x_2^2 + x_3^2/3. \quad (8.80)$$

The equation of state w_ϕ and the density parameter Ω_ϕ of the scalar field are

$$w_\phi = \frac{x_1^2 - x_2^2}{x_1^2 + x_2^2}, \quad \Omega_\phi = x_1^2 + x_2^2. \quad (8.81)$$

Note that from Eq. (8.76) we obtain the relation $\Omega_m = 1 - x_1^2 - x_2^2 - x_3^2$.

The autonomous equations for x_1 , x_2 , and x_3 are given by

$$\frac{dx_1}{dN} = -3x_1 + \frac{\sqrt{6}}{2}\lambda x_2^2 - x_1 \frac{1}{H} \frac{dH}{dN} - \frac{\sqrt{6}}{2}Q(1 - x_1^2 - x_2^2 - x_3^2), \quad (8.82)$$

$$\frac{dx_2}{dN} = -\frac{\sqrt{6}}{2}\lambda x_1 x_2 - x_2 \frac{1}{H} \frac{dH}{dN}, \quad (8.83)$$

$$\frac{dx_3}{dN} = -2x_3 - x_3 \frac{1}{H} \frac{dH}{dN}. \quad (8.84)$$

There are eight fixed points in total, see Table 8.1. The stability of the fixed points can be analyzed by considering three eigenvalues of the Jacobian matrix of perturbations δx_1 , δx_2 , and δx_3 about each point (see problem 8.3).

Among the eight fixed points presented in Table 8.1, we now identify the points responsible for radiation, matter, and accelerated eras.

- (i) Radiation era

The radiation-dominated epoch can be realized either by the points (e), (f), or (g) because they correspond to $w_{\text{eff}} = 1/3$. However the nucleosynthesis bound (7.50) places the constraint $Q^2 > 3.7$ and $\lambda^2 > 88.9$ for the points (f) and (g), respectively. The former case is not compatible with the presence of the matter-dominated epoch, whereas in the latter case λ is too large to have a late-time accelerated solution (as we will see later). Hence the point (e) is the only plausible radiation solution. The eigenvalues of the 3×3 Jacobian matrix for perturbations about the point (e) are

$$\mu = -1, 1, 2. \quad (8.85)$$

This means that the point (e) is a saddle followed by a matter era.

- (ii) Matter era

The matter-dominated epoch can be realized either by the points (a) or (d). Both (a) and (d) correspond to scaling solutions with constant Ω_ϕ and w_ϕ . The point (a) is called the “ ϕ -matter-dominated epoch (ϕ MDE)” [17]. In order for the ϕ MDE to be responsible for the matter era we require that $Q^2 \ll 1$ from the condition $\Omega_\phi = 2Q^2/3 \ll 1$. The eigenvalues of the Jacobian matrix for perturbations about the point (a) are

$$\mu = \frac{3}{2} + Q(Q + \lambda), \quad -\frac{3}{2} + Q^2, \quad -\frac{1}{2} + Q^2. \quad (8.86)$$

As long as $Q^2 \ll 1$, two of the eigenvalues are negative. One of them is positive for $Q(Q + \lambda) > -3/2$, which is satisfied unless $Q < 0$ and $\lambda \gg 1$. Hence the ϕ MDE is a saddle followed by a late-time accelerated point.

Since the effective equation of state for the point (d) is given by $w_{\text{eff}} = -Q/(Q + \lambda)$, it is possible to have $w_{\text{eff}} \simeq 0$ for $|\lambda| \gg |Q|$. The eigenvalues of the Jacobian matrix for perturbations about the point (d) are

$$\mu = -\frac{4Q + \lambda}{2(Q + \lambda)}, \quad -\frac{3(2Q + \lambda)}{4(Q + \lambda)} \left[1 \pm \sqrt{1 + \frac{8[3 - \lambda(Q + \lambda)][3 + 2Q(Q + \lambda)]}{3(2Q + \lambda)^2}} \right]. \quad (8.87)$$

This means that the point (d) is stable for $|\lambda| \gg |Q|$ (either a stable node or a stable spiral). Hence the solutions do not exit from the matter era to the accelerated epoch.

- (iii) Accelerated era

The late-time cosmic acceleration can be realized either by the point (c) or (d). When $\lambda^2 < 2$ the point (c) satisfies the condition for acceleration. The eigenvalues of the Jacobian matrix of perturbations about the point (c) are

$$\mu = \frac{1}{2}(\lambda^2 - 4), \quad \frac{1}{2}(\lambda^2 - 6), \quad \lambda(Q + \lambda) - 3. \quad (8.88)$$

Under the condition $\lambda^2 < 2$, this point is stable for

$$\lambda(Q + \lambda) < 3. \quad (8.89)$$

The energy fraction of the field for the point (d) satisfies $\Omega_\phi > 1$ under the condition (8.89). For the point (d) the condition for the acceleration, $w_{\text{eff}} < -1/3$, corresponds to $Q > \lambda/2$ or $Q < -\lambda$ (recall that λ is assumed to be positive). In both cases the inside of the root of Eq. (8.87) is larger than unity under the condition (8.89) with $-\frac{4Q+\lambda}{2(Q+\lambda)} < 0$ and $-\frac{3(2Q+\lambda)}{4(Q+\lambda)} < 0$. Hence one of the eigenvalues in Eq. (8.87) is positive, which means that the point (d) is a saddle if the point (c) is stable. If the condition

$$\lambda(Q + \lambda) > 3 \quad (8.90)$$

is satisfied, the point (d) is stable whereas the point (c) is a saddle.

Then the late-time stable accelerated solution can be realized by the point (d) under the conditions (8.90) and $Q > \lambda/2$ or $Q < -\lambda$. The scaling solution (d) allows the interesting possibility of a global accelerated attractor with $\Omega_\phi \simeq 0.7$ [40, 284]. However it is difficult to realize the ϕ MDE solution (a) followed by the scaling solution (d). This comes from the fact that the condition $Q^2 \ll 1$ is required to have a ϕ MDE compatible with observations whereas large values of $|Q|$ are needed to get the late-time cosmic acceleration. One can show that there are no allowed regions in the (Q, λ) plane corresponding to the sequence from the ϕ MDE to the scaling attractor [17]. We require a step-like function of the coupling Q in order to realize two scaling solutions [284].

From the above discussion we find that the following sequence is cosmologically viable:

$$(e) \rightarrow (a) \rightarrow (c). \quad (8.91)$$

The presence of the saddle ϕ MDE demands the conditions $Q^2 \ll 1$ and $Q(Q + \lambda) > -3/2$. The stability of the accelerated point requires the conditions $\lambda^2 < 2$ and $\lambda(Q + \lambda) < 3$.

In Fig. 8.3 we plot the cosmological evolution of the density parameters Ω_ϕ , Ω_m , Ω_r as well as the equations of state w_ϕ and w_{eff} for $\lambda = 0.1$ and $Q = 0.3$. This shows that the matter era is in fact replaced by the ϕ MDE with $\Omega_\phi = w_{\text{eff}} \simeq 2Q^2/3 \simeq 0.06$. The ϕ MDE is followed by the accelerated point (c) with the future asymptotic values of the equations of state: $w_\phi = w_{\text{eff}} = -1 + \lambda^2/3 \simeq -0.996$.

The presence of the ϕ MDE changes the background expansion history of the Universe. Since the evolution of the scale factor during the ϕ MDE is given by $a \propto t^{2/(3+2Q^2)}$, the Hubble parameter evolves as $E(z)/E_0 \simeq [\Omega_m^{(0)}(1+z)^{3+2Q^2}]^{1/2}$. Therefore the sound horizon at the decoupling epoch, defined in Eq. (5.34), is smaller than in the uncoupled case by roughly a factor $z_{\text{dec}}^{Q^2}$. For $Q = 0.1$, for instance, this gives a sound horizon 7% smaller. This is a large effect that can be constrained by current measurements, although it is partially compensated by the fact that the distance to the last scattering increases, that is, the CMB shift parameter $\mathcal{R}$ given in Eq. (5.33) is smaller relative to the case $Q = 0$. A full comparison with

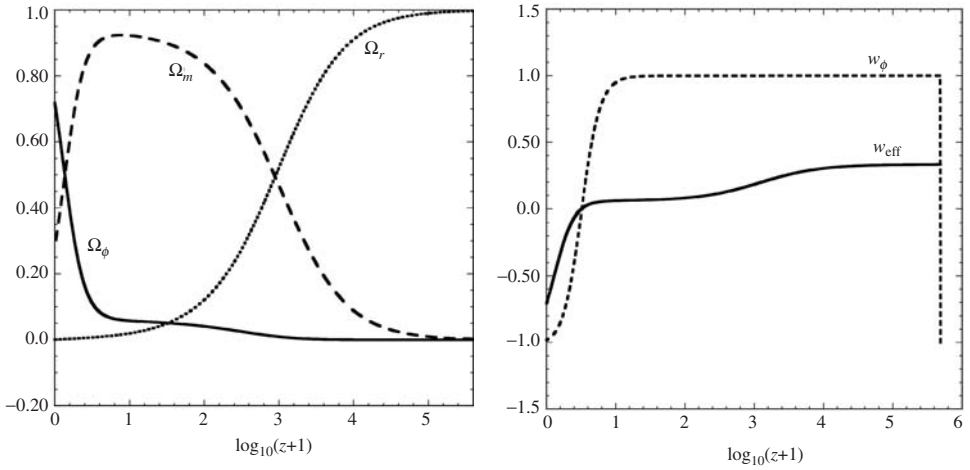


Figure 8.3 Cosmological evolution of the coupled quintessence scenario with an exponential potential for $\lambda = 0.1$ and $Q = 0.3$. The initial conditions are chosen to be $x_1 = 0$, $x_2 = 10^{-10}$, and $x_3 = 0.999$ at the redshift $\log_{10}(z+1) = 5.6919$. The field equation of state w_ϕ starts from -1 because $x_1 = 0$ initially, but it quickly approaches $w_\phi = 1$ due to the dominance of the field kinetic energy relative to the potential energy. This phase is followed by the ϕ MDE (a) in which the potential energy is completely negligible relative to its kinetic energy ($x_1 = -\sqrt{6}Q/3$, $x_2 = 0$). The present epoch is on the way to the accelerated fixed point (c) with $w_{\text{eff}} = w_\phi = -1 + \lambda^2/3$ and $\Omega_\phi = 1$.

CMB data varying also all other parameters shows that the coupling cannot exceed $Q \approx 0.1$ [285].

As we will see later, in coupled quintessence, the equation of matter perturbations is subject to change compared to the uncoupled case. The presence of the coupling between the non-relativistic matter and the scalar field leads to a larger growth rate of matter perturbations relative to the uncoupled quintessence. Hence the observational data of galaxy clustering can be used to place bounds on the strength of the coupling Q . Finally, it is interesting to note that the coupling is partially degenerate with massive neutrinos so that if large neutrino masses are found, as in some laboratory experiment, these can be reconciled with microwave background upper limits [286].

8.3.2 Decoupling the baryons

The scalar field coupling induces a variation of the particle masses. As it can be seen from the conservation equation (8.74) the matter density varies as

$$\rho_m = \rho_m^{(0)} (a/a_0)^{-3} \exp \left(\int_{\phi_0}^{\phi} Q(\tilde{\phi}) d\tilde{\phi} \right), \quad (8.92)$$

where ϕ_0 is the field value today. This can be interpreted at the classical level as a variation of the coupled particle's masses as

$$m = m^{(0)} \exp \left(\int_{\phi_0}^{\phi} Q(\tilde{\phi}) d\tilde{\phi} \right). \quad (8.93)$$

Since in gravitational interactions one always measures the product of masses times the gravitational constant, the limits to the variation of G apply directly to the variation of baryon masses. Current limits on the variation of G are [287]

$$\left| \frac{\dot{G}}{G} \right| \lesssim \text{few} \times 10^{-11} \text{ yrs}^{-1}. \quad (8.94)$$

Then we have, in the case of a constant coupling Q ,

$$\left| \frac{1}{G} \frac{dG}{dN} \right| = \left| \frac{1}{m} \frac{dm}{dN} \right| = \left| Q \frac{d\phi}{dN} \right| \lesssim 10^{-1}. \quad (8.95)$$

Note that we have used the present value of the Hubble parameter given in Eq. (2.36). For instance, on the solution (d) of the previous section, one has $d\phi/dN = \sqrt{6}x_1 = 3/(Q + \lambda)$ so that we find $Q \lesssim (Q + \lambda)/30$ (assuming both $Q, \lambda > 0$). This condition implies that $\lambda \gg Q$, in which case we have no cosmic acceleration ($w_{\text{eff}} = -Q/(Q + \lambda) \approx 0$).

Beside the variation of G , the field coupling Q is constrained by local gravity experiments. There are however several ways to escape these constraints. One, to be discussed later on, is the chameleon mechanism, that screens the effect of the field interaction near high-density objects (as we will see in Section 8.4). Another one is to assume that the coupling varies in time so that it is very small today but large in the past. Yet another solution is to assume that the field couples only to dark matter particles and not to baryons (or couples extremely weakly). In this way, local gravity constraints are emptied of any effect on cosmology [271].

If baryons are uncoupled to the scalar field, their conservation equation is standard and another degree of freedom $x_4 = \sqrt{\rho_b}/(\sqrt{3}H)$ must be added to the dynamical system (8.82)–(8.84) (see Ref. [288]). Since baryons correspond to only a small fraction of the total cosmic fluid today, their effect is in general modest and the fixed points of Table 8.1 remain.

However, there is one case in which the baryons (or in general any uncoupled matter) make a large difference, namely the scaling attractor (d). Scaling solutions are interesting for several reasons. First, they are particularly simple because the density parameters Ω_M, Ω_ϕ and the equation of state are constant. Second, they could help to solve the coincidence problem since the ratio of matter to dark energy is constant not only at the present time but ever since the scaling attractor is reached [288]. Third, they lead to phenomena that cannot be found in non-scaling behaviors,

as for instance an early start of cosmic acceleration. This third point is connected to the uncoupled component, so we discuss it now.

The coupled components, cold dark matter (density ρ_c) and the scalar field (density ρ_ϕ), behave at the background level as a single fluid Ω_{eff} with an effective constant equation of state w_{eff} , i.e. $\rho_c \propto \rho_\phi \propto a^{-3(1+w_{\text{eff}})}$. For the scaling solution (d) we have $w_{\text{eff}} = -Q/(Q + \lambda)$. Let us assume that the condition for the cosmic acceleration is fulfilled. Then since the baryon density evolves as $\rho_b \propto a^{-3}$, baryons dominate in the past for $a < a_b$ where

$$a_b = \left(\Omega_b^{(0)} / \Omega_{\text{eff}}^{(0)} \right)^{-1/(3w_{\text{eff}})}, \quad (8.96)$$

and $\Omega_{\text{eff}}^{(0)} = \Omega_c^{(0)} + \Omega_\phi^{(0)} = 1 - \Omega_b^{(0)}$ (assuming a flat spacetime). Therefore the Universe undergoes a baryon-dominated epoch before a_b in which the expansion and the growth of structure are standard but driven by the baryon rather than by the dark matter. When $a > a_b$ the cosmic expansion is governed by the total effective fluid. If $\Omega_b^{(0)} \approx 0.04$ and $w_{\text{eff}} \approx -1$, one finds $a_b \approx 0.35$ or $z_b \approx 1.9$. Hence we can expect the cosmic expansion to be accelerated from z_b onward. More exactly, one can derive the onset of acceleration (redshift z_{acc}) by solving for $\ddot{a} = 0$ in the flat-space Friedmann equation:

$$\begin{aligned} \frac{\ddot{a}}{a} &= -\frac{4\pi}{3} [\rho_{\text{eff}}(1 + 3w_{\text{eff}}) + \rho_b] \\ &= -\frac{1}{2} H_0^2 \left[\Omega_{\text{eff}}^{(0)} a^{-3(1+w_{\text{eff}})} (1 + 3w_{\text{eff}}) + (1 - \Omega_{\text{eff}}^{(0)}) a^{-3} \right] = 0. \end{aligned} \quad (8.97)$$

This amounts to

$$z_{\text{acc}} = -1 + \left[\frac{\Omega_{\text{eff}}^{(0)} - 1}{\Omega_{\text{eff}}^{(0)}(1 + 3w_{\text{eff}})} \right]^{1/(3w_{\text{eff}})}, \quad (8.98)$$

which gives $z_{\text{acc}} \approx 2.6$ for $\Omega_b^{(0)} = 0.04$ and $w_{\text{eff}} = -1$. This redshift can increase up to $z_{\text{acc}} \approx 4$ for $w_{\text{eff}} = -0.6$, which is actually the value favored by the supernovae data for the scaling case [289]. In general, however, the early acceleration gives a strong integrated Sachs–Wolfe effect on the CMB and to be acceptable it would require other modifications. Notwithstanding these difficulties it is important to derive the conditions for the existence of scaling solutions and we will devote Section 8.5 to this.

8.3.3 Parametrizing coupled dark energy

Let us discuss other coupled dark energy models in which a non-relativistic matter couples to dark energy with an energy density ρ_X and an equation of state w_X . The

interaction between two components can be encoded in the conservation equations:

$$\dot{\rho}_m + 3H\rho_m = +\Gamma\rho_m, \quad (8.99)$$

$$\dot{\rho}_X + 3H(1 + w_X)\rho_X = -\Gamma\rho_m, \quad (8.100)$$

where Γ characterizes the strength of the coupling. The coupled quintessence discussed in the previous subsection corresponds to the choice $\Gamma = Q\dot{\phi}$.

Since the origin of dark energy is not yet identified as a scalar field, we take a different approach to constraining the coupling without assuming scalar fields [274, 275, 276, 277, 278, 279, 280]. We shall measure Γ in terms of the Hubble parameter H and define the dimensionless coupling

$$\delta \equiv \Gamma/H. \quad (8.101)$$

Note that a positive δ corresponds to a transfer of energy from dark energy to dark matter, whereas for a negative δ the energy transfer is opposite. We are interested in placing observational bounds on δ . As usual in the flat Universe the Friedmann equation is given by

$$3H^2 = \rho_m + \rho_r + \rho_X. \quad (8.102)$$

As long as we use cosmic distances whose upper limits of the redshift are smaller than 1000, it is a good approximation to neglect the contribution of radiation.

Equation (8.99) can be written in an integrated form

$$\rho_m = \rho_m^{(0)}(a/a_0)^{-3} \exp\left(\int \delta d(\ln a)\right). \quad (8.103)$$

The cosmological evolution is different depending on the form of the coupling δ . In the following we shall consider two distinct cases.

(A) Constant δ models

For constant δ , Eq. (8.103) is integrated to give

$$\rho_m = \rho_m^{(0)}(a/a_0)^{-3+\delta} = \rho_m^{(0)}(1+z)^{3-\delta}. \quad (8.104)$$

If w_X is constant, substituting Eq. (8.104) into (8.100) leads to the following equation

$$\rho_X = \rho_X^{(0)}(1+z)^{3(1+w_X)} + \rho_m^{(0)} \frac{\delta}{\delta + 3w_X} \left[(1+z)^{3(1+w_X)} - (1+z)^{3-\delta} \right]. \quad (8.105)$$

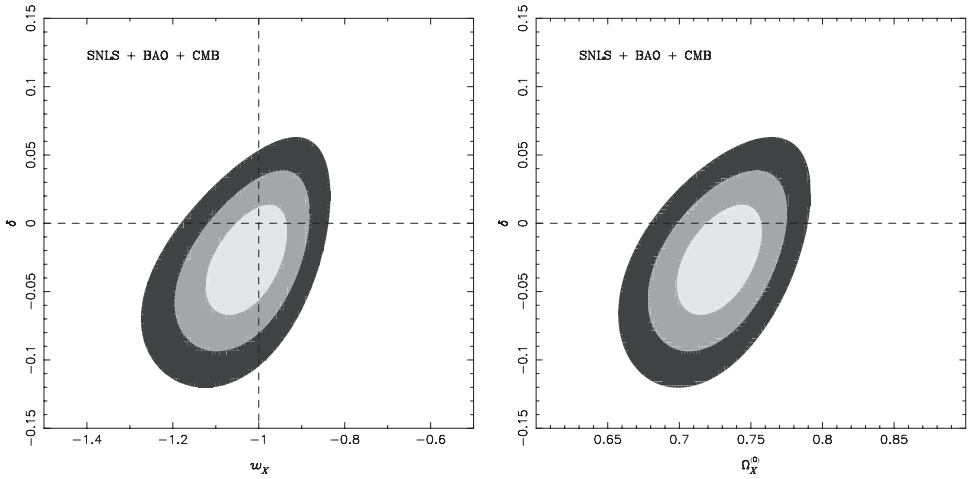


Figure 8.4 Probability contours (1σ , 2σ , and 3σ) for the constant δ parametrization (8.106) obtained from the combined data analysis of SNLS [107], the CMB shift parameter $\mathcal{R}$ [14], and the BAO effective distance D_V [68]. The left panel shows observational contours in the (w_X, δ) plane marginalized over $\Omega_X^{(0)}$, whereas the right panel depicts contours in the $(\Omega_X^{(0)}, \delta)$ plane marginalized over w_X . The Λ CDM model ($w_X = -1$) with no coupling ($\delta = 0$) is in the 1σ contour bound. The best-fit model parameters correspond to $\delta = -0.03$, $w_X = -1.02$, and $\Omega_X^{(0)} = 0.73$. From Ref. [279].

From the Friedmann equation (8.102) and neglecting radiation we obtain

$$E^2(z) = \Omega_X^{(0)}(1+z)^{3(1+w_X)} + \frac{1 - \Omega_X^{(0)}}{\delta + 3w_X} \left[\delta(1+z)^{3(1+w_X)} + 3w_X(1+z)^{3-\delta} \right], \quad (8.106)$$

where $E(z) = H(z)/H_0$ and $\Omega_X^{(0)} = \rho_X^{(0)}/(3H_0^2)$. Hence the Hubble parameter, parametrized in terms of the three parameters $(\delta, w_X, \Omega_X^{(0)})$, is now in a convenient form to confront with the SN Ia, CMB, and BAO observations.

In the high-redshift region ($z \gg 1$), it follows from Eq. (8.105) that ρ_X behaves as $\rho_X \simeq -\rho_m^{(0)} \delta / (\delta + 3w_X)(1+z)^{3-\delta}$ for $3w_X < -\delta$. Hence the energy density of dark energy is negative for $\delta < 0$. We do not exclude such a possibility in the likelihood analysis of model parameters.

Using the parametrization (8.106) it is possible to place observational constraints on the coupling δ from the combined data analysis of the luminosity distance d_L of SN Ia, the CMB shift parameter $\mathcal{R}$, and the BAO effective distance D_V . In Fig. 8.4 we plot observational likelihood contours in both (w_X, δ) and $(\Omega_X^{(0)}, \delta)$ planes for the constant δ model using data from SNLS, CMB, and BAO. The Λ CDM model ($w_X = -1$) with no coupling ($\delta = 0$) is in the 1σ observational contour bound. The allowed observational contours are rather widely spread in the phantom region

($w_X < -1$) with a negative coupling ($\delta < 0$), whereas other parameter regions in the (w_X, δ) plane are also allowed.

(B) Varying δ models

In order to parametrize $H(z)$ for the varying δ case, one way is to assume that the following relation holds between the energy densities ρ_X and ρ_m :

$$\frac{\rho_X}{\rho_m} = \frac{\rho_X^{(0)}}{\rho_m^{(0)}} \left(\frac{a}{a_0} \right)^\xi, \quad (8.107)$$

where ξ is a constant [274]. When $\xi = 0$, Eq. (8.107) corresponds to the scaling relation between dark energy and dark matter. In the absence of the coupling δ with constant w_X the energy density of dark energy evolves as $\rho_X \propto a^{-3(1+w_X)}$, which gives the relation $\xi = -3w_X$. The Λ CDM model corresponds to $w_X = -1$ and $\xi = 3$.

The presence of an interaction between dark energy and dark matter indicates that $\xi \neq -3w_X$. Taking the time derivative of Eq. (8.107), we obtain the relation $\dot{\rho}_X/(H\rho_X) - \dot{\rho}_m/(H\rho_m) = \xi$. Substituting Eqs. (8.99) and (8.100) into this equation, we find that the dimensionless coupling δ can be expressed as

$$\delta(z) = -(\xi + 3w_X)\Omega_X(z), \quad (8.108)$$

where $\Omega_X(z) = \rho_X/(3H^2)$. Since the density parameter $\Omega_X(z)$ is given by $\Omega_X(z) = [(\rho_m^{(0)}/\rho_X^{(0)})(1+z)^\xi + 1]^{-1}$, the coupling $\delta(z)$ varies in time according to

$$\delta(z) = \frac{\delta_0}{\Omega_X^{(0)} + (1 - \Omega_X^{(0)})(1+z)^\xi}, \quad (8.109)$$

where $\delta_0 \equiv -(\xi + 3w_X)\Omega_X^{(0)}$ is the present value of the coupling. When $\xi > 0$ the coupling $|\delta(z)|$ gets smaller for higher z .

Using Eqs. (8.99) and (8.100) together with Eq. (8.107), we obtain the following differential equation for the total density $\rho_T = \rho_m + \rho_X$:

$$\frac{d \ln \rho_T}{da} = -\frac{3}{a} \left[1 + w_X \left\{ \frac{\rho_m^{(0)}}{\rho_X^{(0)}} \left(\frac{a}{a_0} \right)^{-\xi} + 1 \right\}^{-1} \right]. \quad (8.110)$$

This can be integrated to give

$$\rho_T = \rho_T^{(0)} \left(\frac{a}{a_0} \right)^{-3} \left[1 - \frac{\rho_X^{(0)}}{\rho_T^{(0)}} \left\{ 1 - \left(\frac{a}{a_0} \right)^\xi \right\} \right]^{-3w_X/\xi}. \quad (8.111)$$

Hence the evolution of the normalized Hubble parameter is

$$E^2(z) = (1+z)^3 \left[1 - \Omega_X^{(0)} + \Omega_X^{(0)}(1+z)^{-\xi} \right]^{-3w_X/\xi}. \quad (8.112)$$

There are three parameters (ξ , w_X , $\Omega_X^{(0)}$) in this model. Since δ_0 is related to these variables by the relation $\delta_0 = -(\xi + 3w_X)\Omega_X^{(0)}$, one can instead vary three parameters (δ_0 , w_X , $\Omega_X^{(0)}$) when we carry out the likelihood analysis to confront the model with observational data.

The likelihood analysis using the combined data analysis of SNLS [107], the CMB shift parameter $\mathcal{R}$ [14], and the BAO effective distance D_V [68] gives the bounds $-0.4 < \delta_0 < 0.1$, $-1.18 < w_X < -0.91$, and $0.69 < \Omega_X^{(0)} < 0.77$ at the 95% confidence level [279]. Since $\delta(z)$ decreases for larger z , the observational constraint on δ_0 is not so severe compared to the constant coupling model.

From the above results we find that the uncoupled Λ CDM model is well consistent with observational data. The above results are based solely on the modification of the background expansion history of the Universe induced by the coupling. In order to discuss galaxy clustering, for example, we need to study the evolution of matter density perturbations in these coupled dark energy models. In Ref. [290] it was shown that there is an instability of the matter perturbation in the deep radiation-dominated epoch for the constant w_X models. The assumption of constant w_X is restrictive, so one may extend the analysis to more realistic models with varying w_X . While the simple parametrization of the Hubble parameter is difficult in such cases, it will be of interest to pursue the possibility to avoid the instabilities of perturbations.

8.3.4 Mass varying neutrino scenario

It is known that the mass-squared difference between two different species of neutrinos is around $(10^{-2}-10^{-3} \text{ eV})^2$ experimentally, whereas the energy scale of dark energy is about $(10^{-3} \text{ eV})^4$. The similarity between these energy scales has motivated a number of models in which neutrinos and dark energy “know” each other through a direct coupling. In these models, called *mass varying neutrino*, the mass of the neutrino, m_ν , depends on a dark energy scalar field ϕ [291, 292, 293, 294, 295].

Assuming that the distribution of neutrinos is Fermi–Dirac and neglecting its chemical potential, the energy density and the pressure of neutrinos can be expressed, respectively, as [296]

$$\rho_\nu = \frac{T_\nu^4}{\pi^2} \int_0^\infty dy \frac{y^2 \sqrt{y^2 + \xi(\phi)^2}}{e^y + 1}, \quad (8.113)$$

$$P_\nu = \frac{T_\nu^4}{3\pi^2} \int_0^\infty dy \frac{y^4}{\sqrt{y^2 + \xi(\phi)^2} (e^y + 1)}, \quad (8.114)$$

where $\xi(\phi) \equiv m_\nu(\phi)/T_\nu$ and $y \equiv p_\nu/T_\nu$ with T_ν and p_ν being the temperature and the momentum of neutrinos, respectively. We also assume that the scalar field ϕ

has a standard kinetic term with a potential $V(\phi)$. Then the effective potential of the system is given by

$$V_T(\phi) = V(\phi) + (\rho_v - 3P_v). \quad (8.115)$$

The total energy density and the pressure are $\rho_T = \rho_v + \rho_\phi$ and $P_T = P_v + P_\phi$, respectively, where $\rho_\phi = \dot{\phi}^2/2 + V(\phi)$ and $P_\phi = \dot{\phi}^2/2 - V(\phi)$. The total energy satisfies the following conservation law

$$\dot{\rho}_T + 3H(\rho_T + P_T) = 0. \quad (8.116)$$

Substituting the expression of ρ_T into Eq. (8.116), we obtain the equation for ϕ :

$$\ddot{\phi} + 3H\dot{\phi} + V_{,\phi} = -Q(\phi)(\rho_v - 3P_v), \quad Q(\phi) \equiv \frac{d \ln m_v(\phi)}{d\phi}. \quad (8.117)$$

Let us consider the late-time cosmological evolution of the mass varying neutrino scenario in the non-relativistic limit, $m_v \gg T_v$. Note that in the regime where neutrinos are relativistic ($\xi \ll 1$) one has $\rho_v \simeq 3P_v$ so that the r.h.s. of Eq. (8.117) is suppressed. In the non-relativistic limit ($\xi \gg 1$) the neutrino energy density (8.113) and the pressure (8.114) reduce to

$$\rho_v \simeq n_v m_v(\phi), \quad (8.118)$$

$$P_v \simeq 0, \quad (8.119)$$

where n_v is the number density of neutrinos given by

$$n_v = \frac{T_v^3}{\pi^2} \int_0^\infty \frac{dy y^2}{e^y + 1} = \frac{3\zeta(3)}{2\pi^2} T_v^3. \quad (8.120)$$

Hence Eq. (8.117) is approximately given by

$$\ddot{\phi} + 3H\dot{\phi} + V_{,\phi} \simeq -Q(\phi) \rho_v. \quad (8.121)$$

This is the same equation as Eq. (8.77) with $\rho_m = \rho_v$. Thus the mass varying neutrino scenario can be regarded as a coupled quintessence model in which the coupling $Q(\phi)$ depends on the variation of the neutrino mass.

For non-relativistic neutrinos the effective potential (8.115) is given by

$$V_T(\phi) \simeq V(\phi) + n_v m_v(\phi). \quad (8.122)$$

Even if the field potential $V(\phi)$ does not have a minimum, the term $n_v m_v(\phi)$ gives rise to instantaneous minima that vary in time. At the potential minimum we have $\partial V_T / \partial \phi = 0$ and hence

$$n_v = -\frac{\partial V(\phi)}{\partial m_v(\phi)}. \quad (8.123)$$

The potential minimum is present provided that $\partial V/\partial m_\nu < 0$. The kinetic energy of the field ϕ can be neglected relative to its potential energy at the epoch of cosmic acceleration, so that the total energy density and pressure are approximately given by $\rho_T \simeq n_\nu m_\nu + V(\phi)$ and $P_T \simeq -V(\phi)$. Then the equation of state, $w \equiv P_T/\rho_T$, is

$$w = -\frac{V(\phi)}{n_\nu m_\nu + V(\phi)}. \quad (8.124)$$

As long as $n_\nu m_\nu \ll V(\phi)$ it follows that $w \simeq -1$. Hence the mass varying neutrino scenario gives rise to the equation of state satisfying the condition for the late-time cosmic acceleration. The cosmological consequences of the mass varying neutrino scenario such as the effect on CMB and LSS have been extensively studied by a number of authors [297, 298, 299, 300, 301].

A particular case of the mass varying neutrino occurs when the ratio of the neutrino energy density to the scalar field energy density is constant during the accelerated epoch [302]. This is the case of the asymptotic scaling solution (d) discussed in Section 8.3.1 for the exponential potential. Since the effective coupling on the r.h.s. of Eq. (8.117) is suppressed when the neutrino is relativistic but it grows as the neutrino becomes non-relativistic, it is possible that the matter era is followed by the accelerated scaling era (with constant Q asymptotically). If $|Q/\lambda| \gg 1$ one sees that the abundance of the coupled component, in this case neutrinos, is constant and related to the field density by the relation $\Omega_\nu/\Omega_\phi \simeq \Gamma^{-1}$, while the effective equation of state is $w_{\text{eff}} \simeq -1 + \Gamma^{-1}$, where $\Gamma \equiv Q/\lambda$. Therefore the ratio Γ sets at the same time the asymptotic values of the acceleration and of the neutrino density and its mass (averaged over the neutrino families). The fact that the present densities of dark matter and baryons are non-zero implies that we are not yet on the final scaling attractor, so the relations above apply only approximately to the present epoch. In particular, for large Γ we can write $\Omega_\nu \simeq \Omega_\phi/\Gamma \simeq (1 - \Omega_m)/\Gamma$ and therefore $w_{\text{eff}} \simeq -1 + \Omega_\nu/(1 - \Omega_m)$. Since the present energy density of massive neutrinos is $\Omega_\nu^{(0)} \simeq \sum_\nu m_\nu/(94 h^2 \text{ eV})$ (see e.g., [74]), it turns out that the present equation of state in terms of the present neutrino average mass $\hat{m}_\nu \equiv \sum_\nu m_\nu/3$ is given by [302]

$$w_{\text{eff}} \simeq -1 + \frac{(\hat{m}_\nu(t_0)/31 \text{ eV})}{h^2(1 - \Omega_m^{(0)})} \simeq -1 + \frac{\hat{m}_\nu(t_0)}{12 \text{ eV}}. \quad (8.125)$$

In this case observing the equation of state gives information on the neutrino mass, which in turn can be compared with the value derived independently from laboratory measurements.

Beside the equation of state/neutrino mass relation above, the accelerated scaling model makes several definite predictions. Due to the strong coupling required for

the scaling regime to hold, one expects a fast growth of perturbations in the neutrino component. This might lead to interesting phenomena of neutrino-driven clustering and of formation of large neutrino lumps but on the other hand severely limits the viable parametric range of the model [303].

8.3.5 Couplings to the electromagnetic field

Recently there has been some controversial observational evidence in favor of a temporal variation of the effective fine structure constant α . For example, the Oklo natural fission reactor [304] constrained the variation of α relative to the present value α_0 to be $-0.9 \times 10^{-7} < \Delta\alpha/\alpha \equiv (\alpha - \alpha_0)/\alpha_0 < 1.2 \times 10^{-7}$ at the redshift $z \sim 0.16$. The absorption line spectra of distant quasars have given the constraints $\Delta\alpha/\alpha = (-0.574 \pm 0.102) \times 10^{-5}$ over the redshift range $0.2 < z < 3.7$ [305, 306] and $\Delta\alpha/\alpha = (-0.06 \pm 0.06) \times 10^{-5}$ for $0.4 < z < 2.3$ [307]. Although the possibility of systematic errors still remains [308], this can provide an important implication for dark energy.

The variation of α implies the existence of massless or nearly massless fields coupled to gauge fields. Then quintessence or another type of scalar field (such as k-essence) may be responsible for the time variation of α through an interaction between the field ϕ and an electromagnetic field $F_{\mu\nu}$. The Lagrangian density describing such a coupling is given by

$$\mathcal{L}_F(\phi) = -\frac{1}{4}B_F(\phi)F_{\mu\nu}F^{\mu\nu}. \quad (8.126)$$

Originally Bekenstein introduced an exponential form of the coupling: $B_F(\phi) = e^{-\zeta\kappa(\phi-\phi_0)}$ [309], where ζ is a coupling constant and ϕ_0 is the field value today. Provided that $|\zeta\kappa(\phi - \phi_0)| \ll 1$, the coupling $B_F(\phi)$ has the following linear dependence:

$$B_F(\phi) = 1 - \zeta\kappa(\phi - \phi_0). \quad (8.127)$$

This is only one example for the form of the coupling, see Refs. [310, 311, 312, 313, 314, 315, 316] for other choices of the coupling. Since the fine structure “constant” α is inversely proportional to $B_F(\phi)$ this can be expressed as $\alpha = \alpha_0/B_F(\phi)$, where α_0 is the present value. Then the variation of α for the coupling (8.127) is given by

$$\frac{\Delta\alpha}{\alpha} = \frac{\alpha - \alpha_0}{\alpha_0} \simeq \zeta\kappa(\phi - \phi_0). \quad (8.128)$$

Using the constraint $\Delta\alpha/\alpha \simeq -10^{-5}$ around $z = 3$ [306] coming from the information of quasar absorption lines, the coupling ζ can be expressed as

$$\zeta \simeq -\frac{10^{-5}}{\kappa\phi(z=3) - \kappa\phi(z=0)}. \quad (8.129)$$

The ratio of the variation of α around the present can be evaluated as

$$\frac{\dot{\alpha}}{\alpha} \simeq \zeta \kappa \dot{\phi} \simeq -\zeta \frac{d(\kappa\phi)}{d(1+z)} H_0, \quad (8.130)$$

where H_0 is the Hubble parameter today. The bound on $\dot{\alpha}/\alpha$ from atomic clocks is $|\dot{\alpha}/\alpha|_{z=0} < 4.2 \times 10^{-15} \text{ yr}^{-1}$ [317].

As an example, let us consider quintessence with the inverse power-law potential $V(\phi) = M^{4+n} \phi^{-n}$. In the tracking regime of the matter-dominated era, we have shown that the field equation of state is constant: $w_\phi = -2/(n+2)$, see Eq. (7.42). It then follows that the quantity $x = \dot{\phi}^2/(2V)$ defined in Eq. (7.29) is also constant, i.e.

$$\frac{\dot{\phi}^2}{2M^{4+n}\phi^{-n}} = \frac{n}{n+4}. \quad (8.131)$$

Integrating this equation, we find that the evolution of the field in the tracking regime is given by

$$\phi \propto t^{2/(n+2)} \propto a^{3/(n+2)}. \quad (8.132)$$

After the system enters the accelerated epoch, this solution begins to lose its accuracy because the variation of w_ϕ occurs (see Fig. 7.3). Let us proceed to estimate the variation of α under the approximation that the field evolution is given by Eq. (8.132) up to the present epoch, i.e.

$$\phi \approx \phi_0 \left(\frac{a}{a_0} \right)^{3/(n+2)} = \phi_0 (1+z)^{-3/(n+2)}. \quad (8.133)$$

We then find that the coupling ζ in Eq. (8.129) is

$$\zeta \approx \frac{10^{-5}}{1 - 4^{-3/(2+n)}} (\kappa\phi_0)^{-1}. \quad (8.134)$$

Recall that $\kappa\phi_0$ is of the order of unity in order to realize the present cosmic acceleration. Since the denominator in Eq. (8.134) is slightly smaller than 1 for $n = \mathcal{O}(1)$, the coupling ζ is constrained to be $\zeta \approx 10^{-5}$. More careful analysis [315] shows that ζ is in fact of this order for the consistency with the quasar bound.

Using the approximate solution (8.133), the variation of α today can be estimated from Eq. (8.130):

$$\left(\frac{\dot{\alpha}}{\alpha} \right)_{z=0} \approx \frac{3}{n+2} (\kappa\phi_0) \zeta H_0 \approx \zeta H_0, \quad (8.135)$$

where the last approximate equality holds for $n = \mathcal{O}(1)$. Recall that H_0 is given by $H_0 = (h/9.78) \times 10^{-9} \text{ yr}^{-1}$ with $h \approx 0.7$. When $\zeta \approx 10^{-5}$ it follows that $(\dot{\alpha}/\alpha)_{z=0} \approx 10^{-15} \text{ yr}^{-1}$, which satisfies the atomic clock constraints. More stringent constraints, as in Ref. [318], can be escaped by a suitable modulation of the potential [319].

The variation of α for other quintessence potentials was discussed in detail in Ref. [315]. There is also a k-essence model in which a tachyon field is coupled to electromagnetic fields [320]. In this case the form of the coupling $B_F(\phi)$ is naturally determined by underlying particle theory.

8.4 Chameleon scalar fields

If a scalar field ϕ couples to a non-relativistic matter as in the coupled quintessence scenario, this gives rise to a fifth force interaction which can be constrained experimentally. In fact, a coupling Q of the order of unity often arises in superstring and supergravity theories. The existence of such a strongly coupled scalar field is not in general compatible with local gravity experiments unless some mechanism exists to suppress the propagation of the fifth force.

There is an interesting attempt called the *chameleon mechanism* to reconcile large coupling models with local gravity constraints [321, 322]. This is based on a coupled quintessence field whose effective mass is different depending on the environment it is in. The matter coupling gives rise to an extremum of the field potential the field can sit on. If the matter density is sufficiently high as in the interior of a compact object, the field acquires a heavy mass about the potential minimum so that it cannot propagate freely. Meanwhile the field has a lighter mass in a low-density environment such as the exterior of the same compact object.

An effective coupling Q_{eff} between the scalar field ϕ and the non-relativistic matter can be much smaller than its bare coupling Q when a spherically symmetric body has a *thin-shell* around the surface of the body [321, 322]. The field is nearly frozen around the potential extremum at $\phi = \phi_A$ in the region $0 < r < r_1$, where r_1 is close to the radius r_c of the body. In the thin-shell region ($r_1 < r < r_c$) the field begins to evolve because of the dominance of the matter coupling term $Q\rho_A$, where ρ_A is the mean density of the body. As long as the condition $(r_c - r_1)/r_c \ll 1$ is satisfied, it is possible to make the effective coupling $|Q_{\text{eff}}|$ small so that the models are consistent with local gravity experiments. In the following we shall discuss the chameleon mechanism and the resulting field profile in detail.

The action of the chameleon scalar field ϕ with a potential $V(\phi)$ is similar to that of the coupled quintessence scenario:

$$S = \int d^4x \sqrt{-g} \left[\frac{1}{2} R - \frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - V(\phi) \right] - \int d^4x \mathcal{L}_m(g_{\mu\nu}^{(i)}, \Psi_m^{(i)}), \quad (8.136)$$

where g is the determinant of the metric $g_{\mu\nu}$ (in the Einstein frame) and $\mathcal{L}_m$ is a matter Lagrangian with $\Psi_m^{(i)}$ being the matter fields coupled to a metric $g_{\mu\nu}^{(i)}$. The

metric $g_{\mu\nu}^{(i)}$ is related to the Einstein frame metric $g_{\mu\nu}$ via

$$g_{\mu\nu}^{(i)} = e^{2Q_i\phi} g_{\mu\nu}, \quad (8.137)$$

where Q_i are the strengths of the couplings for each matter component with the field ϕ . We use the unit $\kappa^2 = 8\pi G = 1/M_{\text{pl}}^2 = 1$, but we restore M_{pl} or G when it is needed.

The action (8.136) can originate from a theory in which the field ϕ has a direct interaction of the form $e^{-2Q\phi} \tilde{R}$ with the Ricci scalar $\tilde{R}$. This belongs to a class of scalar tensor theory whose action is given by

$$\begin{aligned} \tilde{S} = \int d^4x \sqrt{-\tilde{g}} & \left[\frac{1}{2} e^{-2Q\phi} \tilde{R} - \frac{1}{2} (1 - 6Q^2) e^{-2Q\phi} (\tilde{\nabla}\phi)^2 - U(\phi) \right] \\ & - \int d^4x \mathcal{L}_m(\tilde{g}_{\mu\nu}, \Psi_m). \end{aligned} \quad (8.138)$$

Here a tilde represents quantities in a frame (called the *Jordan frame*) where ϕ has a direct interaction with $\tilde{R}$. As we will see in Section 9.2, the action (8.138) is equivalent to that in Brans–Dicke theory with a potential $U(\phi)$. Under the conformal transformation, $g_{\mu\nu} = e^{-2Q\phi} \tilde{g}_{\mu\nu}$, we obtain the action (8.136) in the Einstein frame, together with the field potential $V(\phi) = U(\phi) e^{4Q\phi}$. Note that the couplings Q_i are the same ($Q_i = Q$) for each matter field. Clearly the metric $g_{\mu\nu}^{(i)}$ in Eq. (8.136) corresponds to the metric $\tilde{g}_{\mu\nu}$ in the Jordan frame.

Varying the Einstein frame action (8.136) with respect to the field ϕ , we get

$$\square\phi - V_{,\phi} = - \sum_i Q_i e^{4Q_i\phi} g_{(i)}^{\mu\nu} T_{\mu\nu}^{(i)}, \quad (8.139)$$

where $T_{\mu\nu}^{(i)} = (2/\sqrt{-g^{(i)}}) \delta\mathcal{L}_m/\delta g_{(i)}^{\mu\nu}$ is the energy-momentum tensor of the i -th matter component. The trace of the i -th component is given by $T^{(i)} \equiv g_{(i)}^{\mu\nu} T_{\mu\nu}^{(i)} = -\tilde{\rho}_i$ for a non-relativistic fluid, where $\tilde{\rho}_i$ is an energy density. If the action (8.136) originates from that in the Jordan frame, then $\tilde{\rho}_i$ represents the energy density in the Jordan frame. We introduce the energy density $\rho_i = \tilde{\rho}_i e^{3Q_i\phi}$ which is conserved in the Einstein frame.² Note that this is different from the energy density $\rho_i^{(E)} = \tilde{\rho}_i e^{4Q_i\phi}$ in the Einstein frame.

Equation (8.139) can be written as

$$\square\phi = V_{,\phi} + \sum_i Q_i \rho_i e^{Q_i\phi}. \quad (8.140)$$

In a spherically symmetric spacetime of the Minkowski background (i.e. weak gravity background) this reduces to

$$\frac{d^2\phi}{dr^2} + \frac{2}{r} \frac{d\phi}{dr} = \frac{dV_{\text{eff}}(\phi)}{d\phi}, \quad (8.141)$$

² In the FLRW cosmological background this means that ρ_i satisfies the equation $\dot{\rho}_i + 3H\rho_i = 0$, while the equation for $\rho_i^{(E)}$ is $\dot{\rho}_i^{(E)} + 3H\rho_i^{(E)} = Q_i\dot{\phi}\rho_i^{(E)}$.

where r is the distance from the center of symmetry and the effective potential V_{eff} is defined by

$$V_{\text{eff}}(\phi) \equiv V(\phi) + \sum_i \rho_i e^{Q_i \phi} . \quad (8.142)$$

In a strong gravity background, Eq. (8.141) is subject to change due to the backreaction of gravitational potentials. In the analysis presented below we focus on the weak gravity background in which the neglect of gravitational potentials can be justified.

We shall consider the case where the couplings Q_i are the same for each matter component, i.e. $Q_i = Q$ together with $\rho = \sum \rho_i$, so that the effective potential is

$$V_{\text{eff}}(\phi) = V(\phi) + \rho e^{Q\phi} . \quad (8.143)$$

We also assume that the spherically symmetric body has a homogeneous density $\rho = \rho_A$ and a mass $M_c = (4\pi/3)\rho_A r_c^3$ with a homogeneous density $\rho = \rho_B$ outside the body. The effective potential V_{eff} has minima at field values ϕ_A and ϕ_B given by

$$V_{,\phi}(\phi_A) + Q\rho_A e^{Q\phi_A} = 0 , \quad (8.144)$$

$$V_{,\phi}(\phi_B) + Q\rho_B e^{Q\phi_B} = 0 . \quad (8.145)$$

The former corresponds to the region with a high density (interior of the body) with a large mass squared $m_A^2 \equiv \frac{d^2 V_{\text{eff}}}{d\phi^2}(\phi_A)$, whereas the latter to the lower density region (exterior of the body) with a small mass squared $m_B^2 \equiv \frac{d^2 V_{\text{eff}}}{d\phi^2}(\phi_B)$. See Fig. 8.5 for illustration.

8.4.1 The field profile of the chameleon field

Equation (8.141) shows that we need to consider the potential $(-V_{\text{eff}})$ in order to find the “dynamics” of the field with respect to r . This means that the effective potential $(-V_{\text{eff}})$ has a maximum at $\phi = \phi_A$, see Fig. 8.6. We impose the following boundary conditions:

$$\frac{d\phi}{dr}(r=0) = 0 , \quad \phi(r \rightarrow \infty) = \phi_B . \quad (8.146)$$

The field ϕ is at rest at $r = 0$ and begins to roll down the potential when the matter-coupling term $Q\rho_A e^{Q\phi}$ becomes important at a radius r_1 . If the field value at $r = 0$ is close to ϕ_A , the field is nearly frozen around ϕ_A in the region $0 < r < r_1$. The body has a thin-shell if r_1 is close to the radius r_c of the body.

In the region $0 < r < r_1$, the r.h.s. of Eq. (8.141) can be approximated as $dV_{\text{eff}}/d\phi \simeq m_A^2(\phi - \phi_A)$ around $\phi = \phi_A$. Hence the solution to Eq. (8.141) is given by $\phi(r) = \phi_A + Ae^{-m_A r}/r + Be^{m_A r}/r$, where A and B are integration constants.

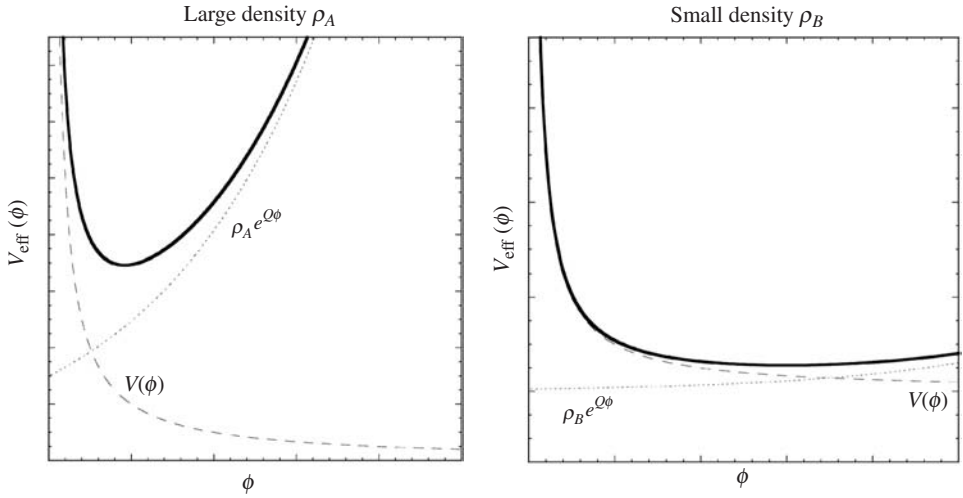


Figure 8.5 The effective potential V_{eff} of a chameleon field (solid curve) for the case $V_{,\phi} < 0$ and $Q > 0$. The effective potential is the sum of the field potential $V(\phi)$ (dashed curve) and the coupling term $\rho e^{Q\phi}$ (dotted curve). The left and right panels correspond to large and small densities, respectively. The field tends to be more massive around the minimum of the effective potential for larger density ρ .

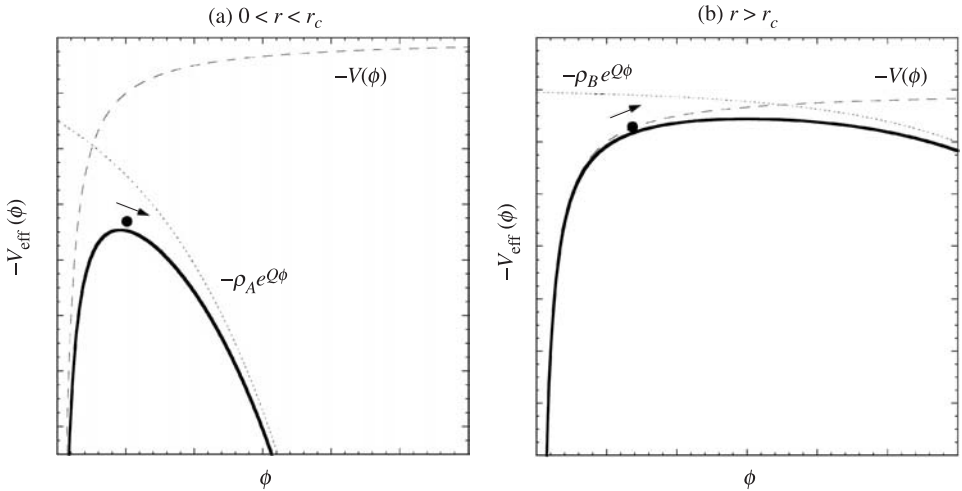


Figure 8.6 The inverted effective potential ($-V_{\text{eff}}$) of a chameleon field inside (left) and outside (right) a spherically symmetric body with constant matter densities ρ_A and ρ_B , respectively. The black points represent the position of the field and show how the field “evolves” with the increase of r . If the body has a thin-shell, the field is nearly frozen in the region $0 < r < r_1$ with $(r_c - r_1)/r_c \ll 1$. The field rolls down the potential for $r_1 < r < r_c$ and it rolls up the potential for $r > r_c$ after acquiring a sufficient kinetic energy in the thin-shell regime ($r_1 < r < r_c$).

Since $B = -A$ to avoid the divergence of ϕ at $r = 0$, the solution reduces to

$$\phi(r) = \phi_A + \frac{A(e^{-m_A r} - e^{m_A r})}{r} \quad (0 < r < r_1), \quad (8.147)$$

which satisfies the boundary condition $\frac{d\phi}{dr}(r = 0) = 0$.

In the region $r_1 < r < r_c$ the field $|\phi(r)|$ evolves toward larger values with the increase of r . In this regime the condition $|V_{,\phi}| \ll |Q\rho_A e^{Q\phi}|$ is satisfied. Since $Q\phi \ll 1$ for most of the field potentials relevant to dark energy, one has $dV_{\text{eff}}/d\phi \simeq Q\rho_A$ in Eq. (8.141). We then find the following solution

$$\phi(r) = \frac{1}{6}Q\rho_A r^2 - \frac{C}{r} + D \quad (r_1 < r < r_c), \quad (8.148)$$

where C and D are constants.

In the region $r > r_c$ the field $|\phi|$ climbs up the potential hill toward larger values. As long as the field acquires a sufficient kinetic energy in the thin-shell regime, the l.h.s. of Eq. (8.141) dominates over the r.h.s. of it. Then the solution outside the body satisfying the boundary condition $\phi(r \rightarrow \infty) = \phi_B$ is given by

$$\phi(r) = \phi_B + \frac{E}{r} \quad (r > r_c). \quad (8.149)$$

If we take into account a small mass term m_B , the solution is given by $\phi(r) \simeq \phi_B + Ee^{-m_B(r-r_c)}/r$. In the following we neglect the contribution of the mass m_B as it does not affect the essential part of the discussion.

The coefficients A , C , D , and E are determined by imposing continuity conditions of $\phi(r)$ and $\phi'(r)$ for the three solutions (8.147), (8.148), and (8.149) at $r = r_1$ and $r = r_c$. We then obtain the following field profile [323]

$$\phi(r) = \phi_A - \frac{1}{m_A(e^{-m_A r_1} + e^{m_A r_1})} \left[\phi_B - \phi_A + \frac{1}{2}Q\rho_A(r_1^2 - r_c^2) \right] \frac{e^{-m_A r} - e^{m_A r}}{r}, \quad (0 < r < r_1), \quad (8.150)$$

$$\begin{aligned} \phi(r) = & \phi_B + \frac{1}{6}Q\rho_A(r^2 - 3r_c^2) + \frac{Q\rho_A r_1^3}{3r} \\ & - \left[1 + \frac{e^{-m_A r_1} - e^{m_A r_1}}{m_A r_1(e^{-m_A r_1} + e^{m_A r_1})} \right] \left[\phi_B - \phi_A + \frac{1}{2}Q\rho_A(r_1^2 - r_c^2) \right] \frac{r_1}{r}, \\ & (r_1 < r < r_c), \end{aligned} \quad (8.151)$$

$$\begin{aligned} \phi(r) = & \phi_B - \left[r_1(\phi_B - \phi_A) + \frac{1}{6}Q\rho_A r_c^3 \left(2 + \frac{r_1}{r_c} \right) \left(1 - \frac{r_1}{r_c} \right)^2 \right. \\ & \left. + \frac{e^{-m_A r_1} - e^{m_A r_1}}{m_A(e^{-m_A r_1} + e^{m_A r_1})} \left\{ \phi_B - \phi_A + \frac{1}{2}Q\rho_A(r_1^2 - r_c^2) \right\} \right] \frac{1}{r}, \\ & (r > r_c). \end{aligned} \quad (8.152)$$

In the original papers of the chameleon mechanism [321, 322], Khoury and Weltman matched two solutions at $r = r_c$ by assuming that the field is frozen in the regime $0 < r < r_1$. In Ref. [323] this was revisited to match three solutions (8.147), (8.148), and (8.149) at $r = r_1$ and $r = r_c$. The radius r_1 is determined by the condition $m_A^2 [\phi(r_1) - \phi_A] = Q\rho_A$, which translates into

$$\phi_B - \phi_A + \frac{1}{2}Q\rho_A(r_1^2 - r_c^2) = \frac{6Q\Phi_c}{(m_A r_c)^2} \frac{m_A r_1 (e^{m_A r_1} + e^{-m_A r_1})}{e^{m_A r_1} - e^{-m_A r_1}}, \quad (8.153)$$

where $\Phi_c = M_c/(8\pi r_c) = \rho_A r_c^2/6$ is the gravitational potential at the surface of the body. Using this relation, the field profile in the region $r > r_c$ is

$$\phi(r) = \phi_B - 2Q_{\text{eff}} \frac{GM_c}{r}, \quad (8.154)$$

where

$$Q_{\text{eff}} = Q \left[1 - \frac{r_1^3}{r_c^3} + 3 \frac{r_1}{r_c} \frac{1}{(m_A r_c)^2} \left\{ \frac{m_A r_1 (e^{m_A r_1} + e^{-m_A r_1})}{e^{m_A r_1} - e^{-m_A r_1}} - 1 \right\} \right]. \quad (8.155)$$

In Refs. [321, 324] the field equation (8.141) was numerically integrated in the case of an inverse power-law potential $V(\phi) = M^{4+n}\phi^{-n}$ by using the solution (8.150) as a boundary condition around $r = 0$. In Ref. [324] it was found that the approximation $dV_{\text{eff}}/d\phi \simeq Q\rho_A$ in the region $r_1 < r < r_c$ overestimates the field value outside the body (typically about 20%). Note that taking into account the mass term m_B outside the body does not change the situation. However, since the order of Q_{eff} does not change, it is a good approximation to use the field profile (8.154) with (8.155) at least for estimating the strength of the effective coupling.

The fifth force on a test particle of unit mass and a coupling Q is given by $F_\phi = -Q\nabla\phi$. Hence its amplitude in the region $r > r_c$ is

$$F_\phi = 2|QQ_{\text{eff}}| \frac{GM_c}{r^2}. \quad (8.156)$$

As long as $|Q_{\text{eff}}| \ll 1$ it is possible to make the fifth force suppressed relative to the gravitational force GM_c/r^2 . From Eq. (8.155) the amplitude of the effective coupling can be made much smaller than $|Q|$ provided that the conditions $\Delta r_c \equiv r_c - r_1 \ll r_c$ and $m_A r_c \gg 1$ are satisfied. Hence we require that the body has a thin-shell and that the field is heavy inside the body for the chameleon mechanism to work.

When the body has a thin-shell ($\Delta r_c \ll r_c$), one can expand Eq. (8.153) in terms of the small parameters $\Delta r_c/r_c$ and $1/(m_A r_c)$:

$$\epsilon_{\text{th}} \simeq \frac{\Delta r_c}{r_c} + \frac{1}{m_A r_c}, \quad (8.157)$$

where ϵ_{th} is the so-called thin-shell parameter defined by

$$\epsilon_{\text{th}} \equiv \frac{\phi_B - \phi_A}{6Q\Phi_c}. \quad (8.158)$$

As long as $m_A r_c \gg (\Delta r_c / r_c)^{-1}$, this recovers the relation $\epsilon_{\text{th}} \simeq \Delta r_c / r_c$ [321, 322]. The effective coupling (8.155) is approximately given by

$$Q_{\text{eff}} \simeq 3Q\epsilon_{\text{th}}. \quad (8.159)$$

Under the condition $\epsilon_{\text{th}} \ll 1$ one has $|Q_{\text{eff}}| \ll |Q|$, which means that the smallness of the thin-shell parameter is crucially important for the compatibility with local gravity constraints. From Eqs. (8.150) and (8.151) the field values at $r = 0$ and $r = r_1$ are $\phi(0) \simeq \phi_A + 12Q\Phi_c / (m_A r_c e^{m_A r_c})$ and $\phi(r_1) \simeq \phi_A + 6Q\Phi_c / (m_A r_c)^2$. This shows that, under the condition $m_A r_c \gg 1$, the field needs to be very close to ϕ_A inside the body to avoid that the field rapidly rolls down the potential because of the heavy mass.

If the field value at $r = 0$ is away from ϕ_A , it begins to roll down the potential at $r = 0$. This is the “thick-shell” solution, which corresponds to taking the limit $r_1 \rightarrow 0$ in Eq. (8.155). One has $Q_{\text{eff}} \simeq Q$ in this limit, so that the model does not satisfy local gravity constraints for values of Q of the order of unity. Hence the body needs to have a thin-shell for the chameleon mechanism to work.

8.4.2 The violation of the equivalence principle

When the spherically symmetric body has a thin-shell, we can place experimental bounds on the thin-shell parameter ϵ_{th} from the possible violation of the equivalence principle (EP). The tightest bound comes from the solar system tests of the weak EP using the free-fall acceleration of the Moon (a_{Moon}) and Earth ($a_{\oplus}$) toward the Sun [322]. The experimental bound on the difference of two accelerations is given by [325]

$$2 \frac{|a_{\text{Moon}} - a_{\oplus}|}{a_{\text{Moon}} + a_{\oplus}} < 10^{-13}. \quad (8.160)$$

Under the conditions that Earth, Sun, and Moon have thin-shells, the field profiles outside the bodies are given as in Eq. (8.154) with the replacement of corresponding quantities. For the field potential that can satisfy local gravity constraints, the condition $|\phi_B| \gg |\phi_A|$ holds for these objects. Then the thin-shell parameters are $\epsilon_{\text{th},\oplus} \simeq \phi_B / (6Q\Phi_{\oplus})$, $\epsilon_{\text{th},\odot} \simeq \phi_B / (6Q\Phi_{\odot})$, $\epsilon_{\text{th},\text{Moon}} \simeq \phi_B / (6Q\Phi_{\text{Moon}})$ for Earth, Sun, and Moon, respectively, with the same asymptotic field value ϕ_B for each object. Note that the gravitational potentials are $\Phi_{\oplus} \simeq 7.0 \times 10^{-10}$, $\Phi_{\odot} \simeq 2.1 \times 10^{-6}$, and $\Phi_{\text{Moon}} \simeq 3.1 \times 10^{-11}$, respectively. Hence we obtain the

following relations

$$\epsilon_{\text{th},\odot} \simeq \frac{\Phi_{\oplus}}{\Phi_{\odot}} \epsilon_{\text{th},\oplus}, \quad \epsilon_{\text{th},\text{Moon}} \simeq \frac{\Phi_{\oplus}}{\Phi_{\text{Moon}}} \epsilon_{\text{th},\oplus}. \quad (8.161)$$

The acceleration induced by a fifth force with the field profile $\phi(r)$ and the effective coupling Q_{eff} is $a^{\text{fifth}} = |Q_{\text{eff}} \nabla \phi(r)|$. Using the thin-shell solution (8.154) for each object, the accelerations $a_{\oplus}$ and a_{Moon} toward the Sun (mass $M_{\odot}$) are [322]

$$a_{\oplus} = \frac{GM_{\odot}}{r^2} (1 + 18Q^2 \epsilon_{\text{th},\oplus} \epsilon_{\text{th},\odot}) \simeq \frac{GM_{\odot}}{r^2} \left(1 + 18Q^2 \epsilon_{\text{th},\oplus}^2 \frac{\Phi_{\oplus}}{\Phi_{\odot}} \right), \quad (8.162)$$

$$a_{\text{Moon}} = \frac{GM_{\odot}}{r^2} (1 + 18Q^2 \epsilon_{\text{th},\text{Moon}} \epsilon_{\text{th},\odot}) \simeq \frac{GM_{\odot}}{r^2} \left(1 + 18Q^2 \epsilon_{\text{th},\oplus}^2 \frac{\Phi_{\oplus}^2}{\Phi_{\odot} \Phi_{\text{Moon}}} \right), \quad (8.163)$$

where we have used Eq. (8.161). Then the condition (8.160) translates into

$$\epsilon_{\text{th},\oplus} < \frac{8.8 \times 10^{-7}}{|Q|}. \quad (8.164)$$

The constraint coming from the violation of strong EP provides a bound $\epsilon_{\text{th},\oplus} \lesssim 10^{-4}$ [322], which is weaker than (8.164) for $|Q| = \mathcal{O}(1)$.

8.4.3 Solar system constraints

In Chapter 9 we will place constraints on modified gravity models by starting from the Jordan frame action given in Eq. (8.138). One can derive the spherically symmetric metric in the Jordan frame by using the technique of the conformal transformation. This is useful to place solar system constraints on modified gravity models by using the so-called post-Newtonian parameter. The spherically symmetric metric in the Einstein frame is given by

$$ds^2 = g_{\mu\nu} dx^{\mu} dx^{\nu} = -[1 - 2\mathcal{A}(r)]dt^2 + [1 + 2\mathcal{B}(r)]dr^2 + r^2 d\Omega^2, \quad (8.165)$$

where $\mathcal{A}(r)$ and $\mathcal{B}(r)$ are functions of the distance r from the center of symmetry and $d\Omega^2 = d\theta^2 + \sin^2 \theta d\phi^2$. The metric outside the spherically symmetric body with mass M_c is given by $\mathcal{A}(r) \simeq \mathcal{B}(r) \simeq GM_c/r$ for weak gravity.

In order to derive the modification induced by the coupling Q we shall consider the metric $g_{\mu\nu}^{(i)}$ defined by Eq. (8.137) or, equivalently, the Jordan frame metric $\tilde{g}_{\mu\nu} = e^{2Q\phi} g_{\mu\nu}$. The line element in the Jordan frame is

$$d\tilde{s}^2 = e^{2Q\phi} ds^2 \quad (8.166)$$

$$= -[1 - 2\tilde{\mathcal{A}}(\tilde{r})]d\tilde{t}^2 + [1 + 2\tilde{\mathcal{B}}(\tilde{r})]d\tilde{r}^2 + \tilde{r}^2 d\Omega^2. \quad (8.167)$$

In the following we shall consider the case $|Q\phi| \ll 1$ and take only the linear terms in $Q\phi$. The condition $|Q\phi| \ll 1$ is satisfied for most of the models we will discuss later. Then we obtain the following relations

$$\tilde{r} = e^{Q\phi} r, \quad \tilde{\mathcal{A}}(\tilde{r}) \simeq \mathcal{A}(r) - Q\phi(r), \quad \tilde{\mathcal{B}}(\tilde{r}) \simeq \mathcal{B}(r) - Qr \frac{d\phi(r)}{dr}. \quad (8.168)$$

Provided that $|Q\phi| \ll 1$ we have that $\tilde{r} \simeq r$. Using the thin-shell solution (8.154) with (8.159), we find that the metrics $\tilde{\mathcal{A}}(r)$ and $\tilde{\mathcal{B}}(r)$ in the Jordan frame are

$$\tilde{\mathcal{A}}(r) = \frac{GM_c}{r} [1 + 6Q^2\epsilon_{\text{th}}(1 - r/r_c)], \quad \tilde{\mathcal{B}}(r) = \frac{GM_c}{r} (1 - 6Q^2\epsilon_{\text{th}}). \quad (8.169)$$

Here we have used the approximation $|\phi_B| \gg |\phi_A|$ and hence $\phi_B \simeq 6Q\Phi_{c\epsilon_{\text{th}}}$. The requirement that the term $Q\phi_B$ does not exceed $\mathcal{A}(r) = GM_c/r$ in Eq. (8.168) gives the condition $r/r_c < (6Q^2\epsilon_{\text{th}})^{-1}$. As long as the field ϕ reaches the value ϕ_B with the distance r_B satisfying the condition $r_B/r_c < (6Q^2\epsilon_{\text{th}})^{-1}$, the metric $\tilde{\mathcal{A}}(r)$ does not change its sign for $r < r_B$. The post-Newtonian parameter γ is given by

$$\gamma \equiv \frac{\tilde{\mathcal{B}}(r)}{\tilde{\mathcal{A}}(r)} \simeq \frac{1 - 6Q^2\epsilon_{\text{th}}}{1 + 6Q^2\epsilon_{\text{th}}(1 - r/r_c)}. \quad (8.170)$$

The present tightest solar system constraint on the post-Newtonian parameter comes from the time-delay effect of the Cassini tracking for the Sun [326]:

$$|\gamma - 1| < 2.3 \times 10^{-5}. \quad (8.171)$$

Since $\gamma \simeq 1 - 6Q^2\epsilon_{\text{th}}$ for $r \approx r_c$ the constraint (8.171) is satisfied for $\epsilon_{\text{th},\odot} < 3.8 \times 10^{-6}/Q^2$, where $\epsilon_{\text{th},\odot}$ is the thin-shell parameter for Sun. Even when $|Q| = \mathcal{O}(1)$, as long as the thin-shell parameter is much smaller than unity the models can be compatible with solar system tests.

The bound (8.171) coming from the solar system constraint is typically weaker than the constraint (8.164) coming from the violation of weak equivalence principle, so we shall adopt the bound (8.164) in the following discussions.

8.4.4 Constraints on the model with an inverse power-law potential

Let us consider experimental bounds on model parameters for the inverse power-law potential

$$V(\phi) = M^{4+n}\phi^{-n}. \quad (8.172)$$

In this subsection we shall consider the positive coupling $Q > 0$ with $\phi > 0$. This is a quintessence potential giving a tracking cosmological solution [38]. In the previous section we have shown that the effective coupling Q_{eff} can be made much smaller than Q for thin-shell solutions with $m_A r_c \gg 1$. Using Eq. (8.158) and the value $\Phi_{\oplus} \simeq 7.0 \times 10^{-10}$ for Earth, the bound (8.164) translates into

$$|\phi_{B,\oplus}| < 3.7 \times 10^{-15}. \quad (8.173)$$

Note that we have employed the relation $|\phi_{B,\oplus}| \gg |\phi_{A,\oplus}|$.

We recall that $\phi_{B,\oplus}$ depends upon the density ρ_B outside Earth. We take the value $\rho_B \simeq 10^{-24} \text{ g/cm}^3$ that corresponds to the mean dark matter/baryon density in our galaxy. For the inverse power-law potential (8.172) we have

$$\phi_{B,\oplus} = \left[\frac{n}{Q} \frac{M_{\text{pl}}^4}{\rho_B} \left(\frac{M}{M_{\text{pl}}} \right)^{n+4} \right]^{\frac{1}{n+1}} M_{\text{pl}}, \quad (8.174)$$

where we have recovered M_{pl} .

Using the bound (8.173) with n and Q of the order of unity, we get the following constraint

$$M \lesssim 10^{-\frac{15n+130}{n+4}} M_{\text{pl}}. \quad (8.175)$$

This shows that $M \lesssim 10^{-2} \text{ eV}$ for $n = 1$ and $M \lesssim 10^{-4} \text{ eV}$ for $n = 2$.

The mass squared m_A^2 about the potential minimum at $\phi = \phi_A$ is given by

$$m_A^2 = n(n+1) \left(\frac{n}{Q} \right)^{-\frac{n+2}{n+1}} \left(\frac{\rho_A}{M_{\text{pl}}^4} \right)^{\frac{n+2}{n+1}} \left(\frac{M}{M_{\text{pl}}} \right)^{-\frac{n+4}{n+1}} M_{\text{pl}}^2. \quad (8.176)$$

Multiplying the square of the Earth radius r_c and eliminating M with the use of Eq. (8.174), we obtain

$$(m_A r_c)^2 = 6(n+1) Q \Phi_{\oplus} \left(\frac{\rho_A}{\rho_B} \right)^{\frac{1}{n+1}} \frac{1}{\phi_{B,\oplus}}. \quad (8.177)$$

Since the mean density of the Earth is $\rho_A \approx 10 \text{ g/cm}^3$, the experimental bound (8.173) leads to

$$m_A r_c > 3\sqrt{(n+1)Q} \cdot 10^{\frac{5n+30}{2(n+1)}}. \quad (8.178)$$

When Q is of the order of unity one has $m_A r_c \gtrsim 10^9$ for $n = 1$ and $m_A r_c \gtrsim 10^7$ for $n = 2$. Hence in these cases the mass m_A in fact satisfies the condition $m_A r_c \gg 1$.

From the above discussion it appears that the local gravity constraints can be satisfied for the inverse power-law potential through the chameleon mechanism. If the same potential is responsible for dark energy, we require that the mass M satisfies the condition (7.34). For n of the order of unity, this requirement is not

compatible with the condition (8.175). However, if we consider the potential of the form $V(\phi) = M^4 \exp(M^n/\phi^n)$ [327], we have $V(\phi) \approx M^4 + M^{4+n}\phi^{-n}$ for $\phi \gtrsim M$ so that the late-time cosmic acceleration is realized for $M \approx 10^{-3}$ eV. This mass scale can then be compatible with the bound coming from local gravity constraints. See Refs. [328, 329] for the possibility of experimental detections of the chameleon field.

8.5 Dark energy models with scaling solutions

We have shown in Section 7.2 that scaling solutions are present for the quintessence scalar field with an exponential potential. The ratio of the field density ρ_ϕ to the fluid density ρ_M as well as the field equation of state w_ϕ is a non-zero (finite) constant for scaling solutions:

$$\frac{\rho_\phi}{\rho_M} = \text{constant} (\neq 0), \quad w_\phi = \text{constant}. \quad (8.179)$$

In this model this solution lies at the boundary between acceleration and deceleration. For a scalar field potential shallower than the exponential potential the solutions finally enter a tracking regime.

Scaling solutions can be also found when dark energy is coupled to matter as we have seen in Section 8.3.1. One example is what we have called ϕ MDE which can replace the standard matter era. Another type is represented by the fixed point (d) in Table 8.1 that can lead to the late-time cosmic acceleration. In this section we shall derive a class of the Lagrangian densities having scaling solutions for general k -essence models.

8.5.1 The condition for the existence of scaling solutions

Let us start with the following general action

$$S = \int d^4x \sqrt{-g_M} \left[\frac{1}{2} R + P(\phi, X) \right] + S_m(g_{\mu\nu}, \phi, \Psi_m^{(i)}), \quad (8.180)$$

where g_M is the determinant of the metric, $P(\phi, X)$ is a Lagrangian density which is the function of a scalar field ϕ and its kinetic energy $X = -(1/2)g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi$. For generality we allow for the interaction between the matter fields $\Psi_m^{(i)}$ and the scalar field ϕ . We shall consider the same form of the coupling as given in Eq. (8.71). We assume that the coupling is a function of the field ϕ , i.e. $Q = Q(\phi)$. Note that the unit $\kappa^2 = 8\pi G = 1$ is chosen.

We focus on the solution with a constant equation of state parameter $w_\phi \equiv P(\phi, X)/\rho_\phi$ in the scaling regime, we assume that the Universe is filled only by two components: the scalar field ϕ and the pressureless matter with an equation of

state $w_m = 0$. Recalling that the field energy density is given by $\rho_\varphi = 2X P_{,X} - P$, the energy densities ρ_φ and ρ_m satisfy the following equations of motion in the flat FLRW spacetime:

$$\frac{d\rho_\varphi}{dN} + 3(1 + w_\varphi)\rho_\varphi = -Q(\varphi)\rho_m \frac{d\varphi}{dN}, \quad (8.181)$$

$$\frac{d\rho_m}{dN} + 3(1 + w_m)\rho_m = +Q(\varphi)\rho_m \frac{d\varphi}{dN}, \quad (8.182)$$

where $N = \ln a$. The Friedmann equation is given by

$$3H^2 = \rho_\varphi + \rho_m. \quad (8.183)$$

We define the fractional densities of ρ_φ and ρ_m as

$$\Omega_\varphi \equiv \frac{\rho_\varphi}{3H^2}, \quad \Omega_m \equiv \frac{\rho_m}{3H^2}, \quad (8.184)$$

which satisfy the relation $\Omega_\varphi + \Omega_m = 1$ from Eq. (8.183).

We are looking for asymptotic scaling solutions where both Ω_φ and w_φ are constants. The condition $\rho_\varphi/\rho_m = \text{constant}$ translates into $d(\ln \rho_\varphi)/dN = d(\ln \rho_m)/dN$. From Eqs. (8.181) and (8.182) we obtain the following relations

$$\frac{d\varphi}{dN} = \frac{3\Omega_\varphi}{Q(\varphi)}(w_m - w_\varphi) \propto \frac{1}{Q(\varphi)}, \quad (8.185)$$

and

$$\frac{d \ln P}{dN} = \frac{d \ln \rho_\varphi}{dN} = \frac{d \ln \rho_m}{dN} = -3(1 + w_{\text{eff}}), \quad (8.186)$$

where

$$w_{\text{eff}} \equiv w_m \Omega_m + w_\varphi \Omega_\varphi. \quad (8.187)$$

Since P is a function of φ and X , Eq. (8.186) translates into

$$\frac{\partial \ln P}{\partial \varphi} \frac{d\varphi}{dN} + \frac{\partial \ln P}{\partial \ln X} \frac{d \ln X}{dN} = -3(1 + w_{\text{eff}}). \quad (8.188)$$

From the definition of X , one gets

$$2X = H^2 \left(\frac{d\varphi}{dN} \right)^2 \propto \frac{\rho_\varphi}{Q(\varphi)^2} \propto \frac{P(\varphi, X)}{Q(\varphi)^2}, \quad (8.189)$$

and thus

$$\frac{d \ln X}{dN} = -3(1 + w_{\text{eff}}) - 2 \frac{d \ln Q}{dN}. \quad (8.190)$$

Substituting Eqs. (8.185) and (8.190) into Eq. (8.188), we arrive at the following equation for the Lagrangian density $P(\varphi, X)$:

$$\left[1 + \frac{2}{\lambda} \frac{dQ(\varphi)}{d\varphi} \right] \frac{\partial \ln P}{\partial \ln X} - \frac{1}{\lambda} \frac{\partial \ln P}{\partial \varphi} = 1, \quad (8.191)$$

where

$$\lambda \equiv Q \frac{1 + w_m - \Omega_\varphi(w_m - w_\varphi)}{\Omega_\varphi(w_m - w_\varphi)}. \quad (8.192)$$

The integration of Eq. (8.191) gives [330]

$$P(\varphi, X) = X Q^2(\varphi) g(X Q^2(\varphi) e^{(\lambda/Q)\psi(\varphi)}), \quad (8.193)$$

where g is an arbitrary function and

$$\psi(\varphi) \equiv \int^\varphi Q(\xi) d\xi. \quad (8.194)$$

If Q is a constant, we obtain [241]

$$P(\varphi, X) = X g(Y), \quad Y \equiv X e^{\lambda\varphi}. \quad (8.195)$$

Note that Q^2 is absorbed in the definition of X . This analysis was extended to a more general cosmological background ($H^2 \propto \rho^n$) in Ref. [331] (see also Refs. [332, 333]).

In the following we focus on the case of the constant coupling Q . As an example, let us consider the Lagrangian density of the form

$$P(\varphi, X) = f(X) - V(\varphi). \quad (8.196)$$

Substituting this into Eq. (8.191), it follows that

$$X \frac{df}{dX} - f(X) = -\frac{1}{\lambda} \frac{dV}{d\varphi} - V \equiv C, \quad (8.197)$$

where C is a constant. The integration of this equation gives $f = c_1 X - C$ and $V = c_2 e^{-\lambda\varphi} - C$. Hence the following Lagrangian density has a scaling solution

$$P = c_1 X - c_2 e^{-\lambda\varphi}. \quad (8.198)$$

This corresponds to the quintessence with an exponential potential.

The dilatonic ghost condensate model (8.7) also has a scaling solution because this Lagrangian density corresponds to the choice $g(Y) = -1 + Y/M^4$ in Eq. (8.195). The tachyon field with the Lagrangian density $P(\phi, X) = -V(\phi)\sqrt{1 - 2X}$ has a scaling solution for the potential $V(\phi) \propto \phi^{-2}$, see problem 8.1.

8.5.2 Fixed points of the scaling Lagrangian

Let us derive the fixed points of the scaling Lagrangian density (8.195) in the presence of non-relativistic matter and radiation. We shall focus on the case $\lambda > 0$ in the following discussion. The energy density of the field is given by $\rho_\phi = X[g + 2Y(dg/dY)]$. Substituting this into Eq. (8.181) yields

$$\ddot{\phi} + 3HAP_{,X\dot{\phi}} + \lambda X[1 - A(g + 2g_1)] = -AQ\rho_m, \quad (8.199)$$

where

$$A(Y) \equiv (g + 5g_1 + 2g_2)^{-1}, \quad g_n \equiv Y^n \partial^n g / \partial Y^n. \quad (8.200)$$

The sound speed, c_s , is related to the quantity A via

$$c_s^2 = AP_{,X}. \quad (8.201)$$

From the stability of quantum fluctuations given in Eq. (8.30), we require the conditions $A > 0$ and $P_{,X} > 0$. The Friedmann equation is

$$3H^2 = X(g + 2g_1) + \rho_m + \rho_r. \quad (8.202)$$

We introduce the following variables

$$x \equiv \frac{\dot{\phi}}{\sqrt{6}H}, \quad y \equiv \frac{e^{-\lambda\phi/2}}{\sqrt{3}H}, \quad u \equiv \frac{\sqrt{\rho_r}}{\sqrt{3}H}. \quad (8.203)$$

Then we find that the variable $Y = Xe^{\lambda\phi}$ is expressed as $Y = x^2/y^2$. The density parameter Ω_ϕ and the equation of state w_ϕ of the field are

$$\Omega_\phi = x^2(g + 2g_1), \quad w_\phi = \frac{g}{g + 2g_1}. \quad (8.204)$$

The density parameter of non-relativistic matter is given by $\Omega_m = 1 - \Omega_\phi - u^2$. It is useful to notice the following relations

$$w_\phi \Omega_\phi = gx^2, \quad w_\phi = -1 + \frac{2x^2}{\Omega_\phi} P_{,X}, \quad (8.205)$$

where $P_{,X} = g + g_1$. The field behaves as a phantom ($w_\phi < -1$) for $P_{,X} < 0$ (provided that $\Omega_\phi > 0$).

The variables x , y , and u obey the following autonomous equations

$$\begin{aligned} \frac{dx}{dN} = \frac{3x}{2} \left[1 + gx^2 - 2A(g + g_1) + \frac{1}{3}u^2 \right] \\ + \frac{\sqrt{6}}{2} [A(Q + \lambda)(g + 2g_1)x^2 - \lambda x^2 + QA(u^2 - 1)], \end{aligned} \quad (8.206)$$

$$\frac{dy}{dN} = \frac{y}{2} \left(3 - \sqrt{6}\lambda x + 3gx^2 + u^2 \right), \quad (8.207)$$

$$\frac{du}{dN} = \frac{u}{2}(-1 + 3gx^2 + u^2). \quad (8.208)$$

The effective equation of state of the system is given by

$$w_{\text{eff}} = gx^2 + \frac{1}{3}u^2. \quad (8.209)$$

We shall derive the fixed points of the above system in the absence of the radiation ($u = 0$). From Eq. (8.207) we have two distinct classes of solutions, either for $3 - \sqrt{6}\lambda x + 3gx^2 = 0$ or for $y = 0$. Using Eq. (8.205), the former case corresponds to

$$x = \frac{\sqrt{6}(1 + w_\varphi \Omega_\varphi)}{2\lambda}. \quad (8.210)$$

Substituting this into Eq. (8.206), we obtain

$$(\Omega_\varphi - 1)[(Q + \lambda)w_\varphi \Omega_\varphi + Q] = 0. \quad (8.211)$$

This allows the following two fixed points *independent of the form of $g(Y)$* [334].

- **(A) Point A: Scalar-field dominated solution:** $\Omega_\varphi = 1$

Substituting $\Omega_\varphi = 1$ into Eqs. (8.205) and (8.210), we get the following relation at the fixed point x_A :

$$w_\varphi = -1 + \frac{\sqrt{6}\lambda x_A}{3} = -1 + \frac{\lambda^2}{3P_{,X}}, \quad (8.212)$$

where $x_A = \lambda/(\sqrt{6}P_{,X})$. In the case of coupled quintessence with an exponential potential ($P = X - V_0 e^{-\lambda\phi}$) one has $w_\varphi = -1 + \lambda^2/3$, which agrees with the equation of state for the fixed point (c) in Section 7.2. The eigenvalues for the Jacobian matrix of perturbations about the point A are given by [334] (see problem 8.4)

$$\mu_+ = -3 + \sqrt{6}(Q + \lambda)x_A, \quad \mu_- = -3 + \frac{\sqrt{6}}{2}\lambda x_A. \quad (8.213)$$

When $Q > -\lambda/2$ and $Q \leq -\lambda/2$ the point A is stable for $x_A < \sqrt{6}/(2(Q + \lambda))$ and $x_A < \sqrt{6}/\lambda$, respectively. For a phantom field ($P_{,X} < 0$) one has $x_A = \lambda/(\sqrt{6}P_{,X}) < 0$

so that the phantom fixed point A is classically stable for $Q > -\lambda$. For a non-phantom field ($x_A > 0$) the point A is stable for

$$P_{,X} > \lambda(Q + \lambda)/3, \quad \text{if } Q > -\lambda/2, \quad (8.214)$$

$$P_{,X} > \lambda^2/6, \quad \text{if } Q \leq -\lambda/2. \quad (8.215)$$

- **(B) Point B: Scaling solution:** $\Omega_\varphi = -Q/(w_\varphi(Q + \lambda))$

Substituting the relation $\Omega_\varphi = -Q/(w_\varphi(Q + \lambda))$ into Eq. (8.210), we find

$$x_B = \frac{\sqrt{6}}{2(Q + \lambda)}. \quad (8.216)$$

We also obtain the following relations

$$w_{\text{eff}} = -\frac{Q}{Q + \lambda}, \quad w_\varphi = -\frac{Q(Q + \lambda)}{Q(Q + \lambda) + 3(g + g_1)}, \quad \Omega_\varphi = \frac{Q(Q + \lambda) + 3(g + g_1)}{(Q + \lambda)^2}. \quad (8.217)$$

The condition for the cosmic acceleration corresponds to $w_{\text{eff}} < -1/3$, i.e.

$$Q > \lambda/2 \quad \text{or} \quad Q < -\lambda. \quad (8.218)$$

In the presence of the coupling Q it is possible to have an accelerated scaling solution with $\Omega_\varphi = \text{constant}$.

The eigenvalues for the Jacobian matrix of perturbations about the point B can be derived without specifying the form of $g(Y)$ [334]:

$$\mu_\pm = -\frac{3(2Q + \lambda)}{4(Q + \lambda)} \left[1 \pm \sqrt{1 - \frac{8(1 - \Omega_\varphi)(Q + \lambda)^3[\Omega_\varphi(Q + \lambda) + Q]}{3(2Q + \lambda)^2}} A \right], \quad (8.219)$$

where A is defined in Eq. (8.200). Under the conditions (8.218), the point B is stable if

$$-\frac{Q}{Q + \lambda} \leq \Omega_\varphi < 1, \quad \text{and} \quad A > 0. \quad (8.220)$$

The second condition is required to avoid the ultra-violet instability of quantum fluctuations (together with the condition $P_{,X} > 0$). Using Eq. (8.217) we find that the condition $-Q/(Q + \lambda) \leq \Omega_\varphi$ corresponds to $-2Q(Q + \lambda) \leq 3P_{,X}$, which is automatically satisfied for a non-phantom scalar field ($P_{,X} > 0$) under the condition (8.218). The crucial condition for the stability of the point B is $\Omega_\varphi < 1$, i.e.

$$P_{,X} < \lambda(Q + \lambda)/3. \quad (8.221)$$

For a non-phantom point B this condition can be satisfied for $Q > \lambda/2$, but not for $Q < -\lambda$. Hence, if $Q > \lambda/2$ and $\Omega_\varphi < 1$, there exists a stable, accelerated, and non-phantom fixed point B. When $Q > \lambda/2$ the stability condition (8.221) has an opposite inequality to that in Eq. (8.214). Hence, when the point B is stable, the point A is not stable, and vice versa.

Let us next study the second class of fixed points, i.e. $y = 0$. The function $g(Y)$ (where $Y = x^2/y^2$) should be non-singular at $y = 0$ for the existence of such points. From this requirement the function g can be expanded in positive powers of y^2/x^2 , i.e.

$$g = c_0 + \sum_{n>0} c_n (y^2/x^2)^n. \quad (8.222)$$

For the function (8.222) it follows that

$$g_n(y \rightarrow 0) = 0 \quad (n > 0). \quad (8.223)$$

Substituting Eq. (8.222) into Eq. (8.206), we find that the fixed points satisfy

$$\frac{dx}{dN} = \frac{1}{2} \left(3c_0 x + \sqrt{6}Q \right) \left(x^2 - \frac{1}{c_0} \right) = 0. \quad (8.224)$$

If $c_0 \neq 0$ we have the following fixed points.

- **(C) φ MDE point C**

The φ MDE corresponds to the point

$$(x_C, y_C) = \left(-\frac{\sqrt{6}Q}{3c_0}, 0 \right), \quad (8.225)$$

with

$$\Omega_\varphi = w_{\text{eff}} = \frac{2Q^2}{3c_0}, \quad w_\varphi = 1. \quad (8.226)$$

The φ MDE is also a scaling solution giving constant values of Ω_φ and w_φ . When $c_0 > 0$ the requirement of the condition $\Omega_\varphi < 1$ gives the bound

$$|Q| < \sqrt{3c_0/2}. \quad (8.227)$$

The eigenvalues for the matrix of perturbations about the φ MDE are

$$\mu_1 = (Q^2/c_0) - 3/2, \quad \mu_2 = 3/2 + Q(Q + \lambda)/c_0. \quad (8.228)$$

When $c_0 > 0$, μ_1 is negative under the condition (8.227), whereas $\mu_2 > 0$ for the values of Q and λ satisfying Eq. (8.218). This shows that for positive c_0 the φ MDE corresponds to a saddle point for all the relevant cases. When $c_0 < 0$ the φ MDE can be a stable point for $Q(Q + \lambda) > 3|c_0|/2$, but such cases are excluded as we will see later.

- **(D) Purely kinetic point**

This point exists for $c_0 > 0$ and is given by

$$(x_D, y_D) = (\pm 1/\sqrt{c_0}, 0), \quad (8.229)$$

with

$$\Omega_\varphi = 1, \quad w_{\text{eff}} = w_\varphi = 1. \quad (8.230)$$

The eigenvalues for the matrix of perturbations are

$$\mu_1 = 3 \pm \sqrt{6/c_0} Q, \quad \mu_2 = 3 \mp \sqrt{6/c_0} \lambda/2, \quad \text{for } x_D = \pm 1/\sqrt{c_0}, \quad (8.231)$$

where the double sign corresponds to the same order. When $Q > 0$ one of the eigenvalues is at least positive for both signs of x_D , which means that the point D is either an unstable node or a saddle. When $Q < 0$ the point corresponding to $x_D = 1/\sqrt{c_0}$ is stable for $Q < -\sqrt{3c_0}/2$ and $\lambda > \sqrt{6c_0}$, whereas the point corresponding to $x_D = -1/\sqrt{c_0}$ is an unstable node.

In summary the scalar-field dominated point A and the scaling solution B are present for any form of $g(Y)$. Since it is possible to have $w_{\text{eff}} < -1/3$ for these points, they can be used for the late-time acceleration. The solutions choose one of the points as a final attractor depending on the initial conditions. The existence of the points C and D depends on the form of the scalar-field Lagrangian. The quintessence with an exponential potential ($g = 1 - c/Y$) belongs to this class, but the dilatonic ghost condensate model ($g = -1 + cY$) does not. The φ MDE point C is a kind of scaling solution by which the standard matter era can be replaced. The point D corresponds to $\Omega_\varphi = 1$, but it does not lead to the cosmic acceleration. This is viable neither for the matter-dominated epoch nor for the dark energy dominated epoch. We note that the radiation fixed point, $(x, y, u) = (0, 0, 1)$, also exists for the dynamical system (8.206)–(8.208).

8.5.3 Two scaling regimes?

For the coupled quintessence with an exponential potential we have shown in Section 8.3 that the typical cosmological trajectory is a sequence of radiation, φ MDE point C, and the scalar-field dominated point A. If it is possible to realize the φ MDE point C followed by the scaling solution B, this is attractive for alleviating the coincidence problem in the sense that dark energy and dark matter follow the same scaling solution from the end of the radiation era. Unfortunately it was shown in Ref. [17] that the coupled quintessence with an exponential potential does not allow for such cosmological evolution. In the following we shall discuss the possibility of realizing two scaling solutions (the φ MDE point C followed by the point B) for a vast class of the coupled scalar-field Lagrangian density (8.195) having scaling solutions. Recall that the φ MDE exists for the function

$$g = c_0 + \sum_{n>0} c_n Y^{-n}, \quad (8.232)$$

where n are integers. We shall discuss three cases: (i) $c_0 > 0$, (ii) $c_0 < 0$, and (iii) $c_0 = 0$, separately.

(i) $c_0 > 0$

The scaling solution B can be used for the late-time cosmic acceleration provided that $Q > \lambda/2 > 0$ or $Q < -\lambda < 0$. In the former case one has $x_B > 0$ and $x_C < 0$ from Eqs. (8.216) and (8.225), whereas in the latter case $x_B < 0$ and $x_C > 0$. The function g given in Eq. (8.232) is singular at $x = 0$. This suggests that the cosmological trajectory from the point C to the point B is prevented.

To be more concrete let us consider a single power-law function of $g(Y)$:

$$g = c_0 - cY^{-p}, \quad (8.233)$$

where p is not necessarily an integer. In the limit that $x \rightarrow 0$ with a non-zero value of y , the term gx^2y on the r.h.s. of Eq. (8.207) shows a divergence for $p > 1$, whereas the term gx^3 on the r.h.s. of Eq. (8.206) diverges for $p > 3/2$. When $p \neq 1$ we have [330]

$$\left| \frac{dy/dN}{dx/dN} \right|_{x \rightarrow 0} \rightarrow \infty, \quad (8.234)$$

which means that the solutions cannot pass the line $x = 0$ in the (x, y) plane. Since the signs of x_B and x_C are different, it is inevitable to hit the singularity for $p > 1$ if the solutions move from the point C to the point B.

We also note that there is another singularity associated with the divergence of the sound speed c_s . From Eq. (8.201) this happens for $A^{-1} = c_0 - c(p-1)(2p-1)Y^{-p} = 0$, i.e.

$$y = \pm \left(\frac{c_0}{c(p-1)(2p-1)} \right)^{1/(2p)} x. \quad (8.235)$$

For positive c , these lines exist for $p > 1$ or $0 < p \leq 1/2$ but disappear for $1/2 < p \leq 1$.

In Fig. 8.7 we see that cosmological trajectories from the point C to the point B cannot avoid the singularity of the sound speed in addition to the presence of the singularity at $x = 0$. Thus we have shown that the φ MDE is not followed by the scaling attractor B for $p \geq 1$.

When $p = 1$, i.e. for coupled quintessence with an exponential potential, we have neither of the two singularities mentioned above. Hence it is possible to realize the trajectories from the point C to the point B, see Fig. 8.7. However, in this model, it is not possible to satisfy the observational requirement $\Omega_\varphi^{(B)} = 0.7 \pm 0.2$ and $w_{\text{eff}}^{(B)} < -0.6 \pm 0.1$ for the point B, while at the same time satisfying the bound $\Omega_\varphi^{(C)} < 0.2$ for the point C to be consistent with CMB and LSS data [17]. Hence the model with $p = 1$ does not allow an ideal cosmological trajectory.

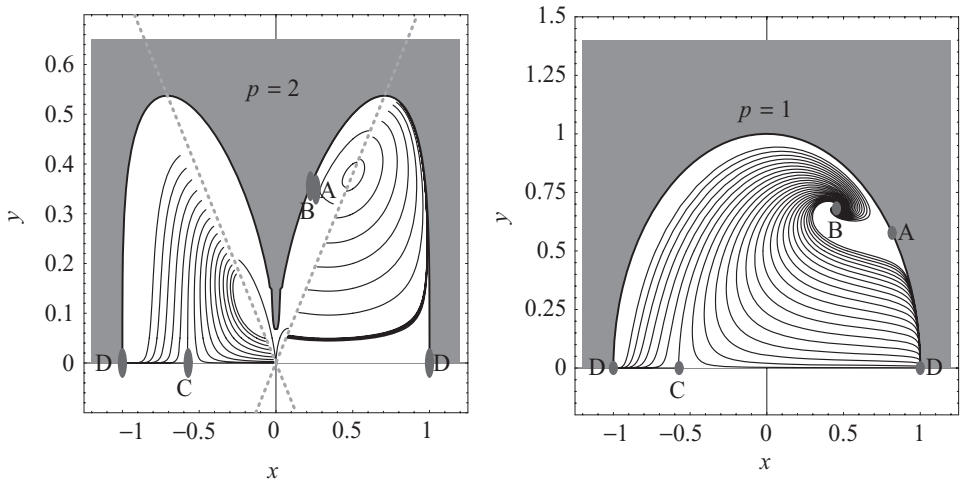


Figure 8.7 Phase space trajectories for the model $g = 1 - Y^{-2}$ with $Q = 0.7$ and $\lambda = 4$ (left) and for the model $g = 1 - Y^{-1}$ (coupled quintessence with an exponential potential) with $Q = 1.02$ and $\lambda = 1.54$ [330]. We also plot the fixed points A, B, C, D. The grey area corresponds to the region $\Omega_\phi > 1$. In the model of the left panel, there is a singularity at $x = 0$ that prevents the cosmological evolution from the point C to the point B. The dotted line represents another singularity given by Eq. (8.235) at which the sound speed diverges. In the model of the right panel the ϕ MDE point C can be followed by the scaling solution B, but such a solution is not cosmologically viable. From Ref. [330].

While we have considered a single power p larger than 1, the result we have obtained above can be applied for the function g of the form (8.232). In fact, if the function g includes any power n larger than 1, this leads to a singularity at $x = 0$ even if the singularity of the sound speed may be avoided. Thus the model (8.232) does not allow the sequence of two scaling solutions for positive c_0 .

For the model (8.233) with $0 < p < 1$ the line $x = 0$ is not singular, but the property (8.234) still holds. Hence it is difficult to connect the point C to the point B. In this case there exists a nearly matter-dominated era with small positive values of x and y , but this is different from the ϕ MDE. So the sequence of two scaling solutions is not realized in this case.

(ii) $c_0 < 0$

When $c_0 < 0$ the ϕ MDE corresponds to a negative value of Ω_ϕ , see Eq. (8.226). Moreover we encounter another problem if we aim to connect the point C to the point B without singularities. The quantity A defined in Eq. (8.200) reduces to $A^{-1} = c_0$ by using the relation (8.223) for the function (8.222). Since $c_0 < 0$ we have that $A < 0$. As we have shown in Eq. (8.220), we require that $A > 0$ to get a viable scaling solution. In order to reach the point B from the point C, we need to cross either $A = 0$ ($x = 0$) or the singularity at $A^{-1} = 0$ (which is not allowed).

The former can be realized only for the model (8.233) with $p = 1$. This discussion shows that the negative c_0 case does not allow the sequence of two scaling solutions.

(iii) $c_0 = 0$

When $c_0 = 0$, Eq. (8.225) shows that there is no φ MDE fixed point and hence this case is not viable either. Although one can have a matter era for $x \rightarrow \infty$ with $n \geq 1$ or for $x \rightarrow 0$ with the power n less than 1, A is singular in both cases.

We have thus shown that no cosmologically viable sequence of two scaling solutions can be realized for a general class of the Lagrangian density with constant coupling Q . We wish to stress here that the above analysis covers all of the scalar-field scaling models proposed in the literature. The point is that we require a small coupling Q during the matter era to keep the condition $\Omega_\varphi \ll 1$, whereas a large coupling is needed to get a sufficient cosmic acceleration at late times. For the constant coupling Q the typical cosmological evolution is the φ MDE point C followed by the accelerated point A. It is possible to have two scaling regimes if the coupling $Q(\varphi)$ changes from a small value Q_1 to a large value Q_2 very rapidly. One such example is

$$Q(\varphi) = \frac{1}{2} \left[(Q_2 - Q_1) \tanh \left(\frac{\varphi - \varphi_1}{\Delta} \right) + Q_2 + Q_1 \right], \quad (8.236)$$

where the growth of the coupling occurs around $\varphi = \varphi_1$ with the variance Δ . In this case the solutions can approach a stationary global attractor with $\Omega_\varphi \simeq 0.7$ preceded by the φ MDE [284]. We note that the mass-varying neutrino scenario discussed in Section 8.3.4 also allows the growth of the effective coupling [302].

8.6 Unified models of dark energy and dark matter

Notwithstanding their radically different properties in terms of the equation of state and clustering, the temptation to unify dark matter and dark energy in a single entity occurred to many cosmologists almost from the beginning. This can be achieved in several ways, with more or less theoretical justification.

In this section we discuss unified models of dark energy and dark matter using a single fluid or a single scalar field [45, 46, 335]. In the following we shall first present the generalized Chaplygin gas (GCG) model as an example of a single fluid model. We then proceed to a class of unified models using a single scalar field.

8.6.1 Generalized Chaplygin gas model

The Chaplygin³ gas model has been proposed by Kamenshchik *et al.* [45]. In this model the pressure P of the perfect fluid is related to its energy density ρ via

³ Named after Sergey Chaplygin (1869–1942), Russian mathematician, physicist, and engineer, who found a similar behavior in aerodynamical studies. Chaplygin is very likely the only scientist who has a lunar crater, a city, and a cosmological model named after him.

$P = -A/\rho$. For generality, one can also work with a generalized Chaplygin gas (GCG) model given by [46]

$$P = -A\rho^{-\alpha}, \quad (8.237)$$

where A is a positive constant. If $\alpha > 0$ the pressure is suppressed relative to the energy density in the early cosmological epoch. At late times the negative pressure becomes important so that the cosmic acceleration can be realized. A fluid with the Chaplygin equation of state therefore interpolates between pressureless matter and dark energy and could in principle replace both.

Plugging the relation (8.237) into the continuity equation (2.20), we obtain the following integrated solution

$$\rho(t) = \left[A + \frac{B}{a^{3(1+\alpha)}} \right]^{1/(1+\alpha)}, \quad (8.238)$$

where B is an integration constant. Here $a(t) = (1+z)^{-1}$ is the scale factor normalized to unity today. From Eq. (8.238) we find that the density ρ evolves as $\rho \propto a^{-3}$ in the early epoch ($a \ll 1$) and $\rho \propto A^{1/(1+\alpha)}$ in the late epoch ($a \gg 1$), respectively. In the flat FLRW background the density ρ is related to the expansion rate H via the relation

$$3H^2 = 8\pi G\rho. \quad (8.239)$$

We introduce the following quantities

$$\rho_* \equiv (A + B)^{1/(1+\alpha)}, \quad \Omega_m^* \equiv \frac{B}{A + B}, \quad (8.240)$$

where Ω_m^* is interpreted as an effective matter density (which is different from $\Omega_m^{(0)}$). Then Eq. (8.238) can be written as

$$\rho(z) = \rho_* \left[1 - \Omega_m^* + \Omega_m^* (1+z)^{3(1+\alpha)} \right]^{1/(1+\alpha)}. \quad (8.241)$$

We then obtain the equation of state

$$w(z) = - \left[1 + \frac{\Omega_m^*}{1 - \Omega_m^*} (1+z)^{3(1+\alpha)} \right]^{-1}. \quad (8.242)$$

In the region of high redshift ($z \gg 1$) this approaches the value $w \approx 0$. The present value of w is $w(0) = -(1 - \Omega_m^*)$. In the asymptotic future the equation of state approaches the value $w \rightarrow -1$.

Thus the GCG model can account for both dark matter and dark energy at least at the background level. Let us consider the evolution of matter density perturbations δ_m in the GCG model [336, 337]. The presence of the pressure gives rise to a modification to the matter perturbation equation given in Eq. (4.77). Each Fourier

mode for the total matter perturbation δ_m^* [defined in Eq. (4.65)] with a comoving wavenumber k obeys Eq. (4.67), i.e.

$$\ddot{\delta}_m^* + (2 + 3c_s^2 - 6w) H \dot{\delta}_m^* - \left[\frac{3}{2} H^2 (1 - 6c_s^2 - 3w^2 + 8w) - \left(\frac{c_s k}{a} \right)^2 \right] \delta_m^* = 0, \quad (8.243)$$

where c_s is the sound speed given by

$$c_s^2 = \frac{\partial P}{\partial \rho} = -\alpha w. \quad (8.244)$$

Since $w \rightarrow 0$ and $c_s^2 \rightarrow 0$ in the limit $z \gg 1$, the sound speed is much smaller than unity in the deep matter era and starts to grow around the end of it. Since $w < 0$ from Eq. (8.242), c_s^2 is positive for $\alpha > 0$ and negative for $\alpha < 0$. From Eq. (8.243) the perturbations satisfying the following condition grow via the gravitational instability [337]:

$$|c_s^2| \lesssim \frac{3}{2} \left(\frac{aH}{k} \right)^2. \quad (8.245)$$

When $c_s^2 \gtrsim \frac{3}{2} (aH/k)^2$, the perturbations exhibit oscillations because of the dominance of the pressure term $(kc_s/a)^2$ relative to the term $3H^2/2$. Using Eq. (8.239) for $c_s^2 > 0$, the condition (8.245) is interpreted as $\lambda_p > \lambda_J$, where $\lambda_p = (2\pi/k)a$ is the physical wavelength and $\lambda_J = |c_s|(\pi/G\rho)^{1/2}$ is the Jeans length. If the condition (8.245) is violated for $c_s^2 < 0$, the perturbations exhibit violent instabilities and grow exponentially. This exponential instability tends to be stronger on smaller scales.

The violation of the condition (8.245) mainly occurs around the present epoch in which $|w|$ is of the order of unity and hence $|c_s^2| \sim |\alpha|$. The typical smallest scale relevant to the galaxy matter power spectrum corresponds to the wavenumber $k = 0.1 h \text{ Mpc}^{-1}$ (as long as the linear perturbations are concerned). Using the value (2.37) for the present Hubble radius, the constraint (8.245) gives [337]

$$|\alpha| \lesssim 10^{-5}. \quad (8.246)$$

In Fig. 8.8 the matter power spectra are plotted for several different values of α . When $\alpha \lesssim -10^{-5}$, the perturbations are exponentially amplified because the sound speed squared is negative. Meanwhile the perturbations exhibit rapid oscillations for $\alpha \gtrsim 10^{-5}$. In both cases the matter power spectra are radically different from that in the Λ CDM model.

The GCG model is therefore not much different from the Λ CDM model if we take the constraint (8.246) at face value.⁴ In particular the original Chaplygin gas

⁴ Note that non-linear clustering may alter the evolution of density perturbations in this model, see Refs. [338, 339]. Moreover the constraints are relaxed if one assumes that beside the Chaplygin gas there is standard dark matter.

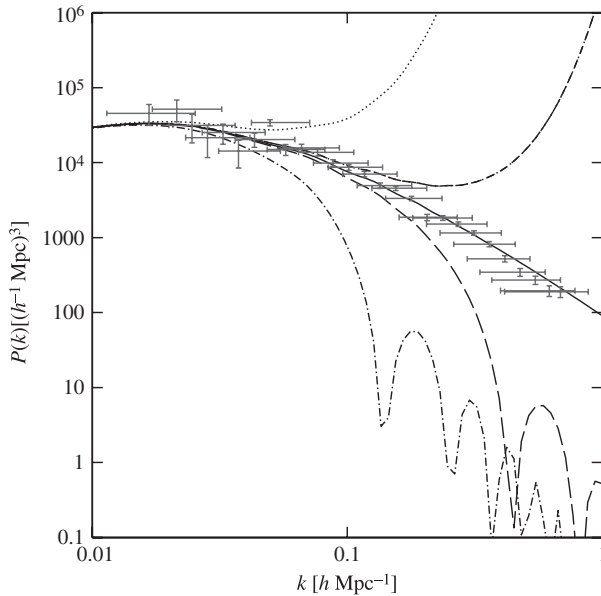


Figure 8.8 The matter power spectra in the GCG model. From top to bottom each curve corresponds to $\alpha = -10^{-4}, -10^{-5}, 0, 10^{-5}, 10^{-4}$, respectively. The data points of the 2dF galaxy redshift survey are also plotted. Compared to the Λ CDM model ($\alpha = 0$) the difference appears for $|\alpha| \gtrsim 10^{-5}$. From Ref. [337].

model ($\alpha = 1$) is completely ruled out from observations. The crucial point is the role of the pressure P in the GCG model. During most of the matter era, the effect of the pressure needs to be strongly suppressed for a successful structure formation (at least in the regime of linear perturbations). However, a negative pressure is required for the late-time cosmic acceleration. In the GCG model it is difficult to satisfy these two demands at the same time.

There are various ways to escape this negative conclusion. One is to avoid the large sound speed problem by adding to the adiabatic sound speed (8.244) a non-adiabatic contribution that makes c_s vanish (silent quartessence [340]). Another one is to consider the Chaplygin gas as a candidate for dark energy alone, instead of a unified fluid. Yet another is to change the equation of state, as in Ref. [341]. A further possibility will be discussed in the next section.

8.6.2 k -essence as unified models of dark energy and dark matter

Let us discuss now the possibility of building unified models using a single field ϕ . One such model using a purely kinetic Lagrangian density is [335]

$$P = F(X), \quad (8.247)$$

where $F(X)$ is a function of the kinetic energy $X = -(1/2)g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi$ of the field ϕ . Since the field energy density and its pressure are given by $\rho_\phi = 2XF_{,X} - F$ and $P_\phi = F$ respectively, the continuity equation (8.15) in the flat FLRW spacetime yields

$$(F_{,X} + 2XF_{,XX})\dot{X} + 6HF_{,X}X = 0. \quad (8.248)$$

This is integrated to give

$$XF_{,X}^2 = Ca^{-6}, \quad (8.249)$$

where C is an integration constant.

Let us now take a function $F(X)$ with an extremum at some value $X = X_0$, such that $F_{,X}(X_0) = 0$ as for instance [335]

$$F(X) = F_0 + F_2(X - X_0)^2, \quad (8.250)$$

where F_0 and F_2 are constants. We shall find the solution of Eq. (8.248) under the condition

$$\epsilon \equiv \frac{X - X_0}{X_0} \ll 1. \quad (8.251)$$

Substituting the function (8.250) into Eq. (8.248), we obtain the following linear-order equation in terms of ϵ :

$$\dot{\epsilon} = -3H\epsilon. \quad (8.252)$$

This gives the following solution

$$X = X_0 [1 + \epsilon_1(a/a_1)^{-3}], \quad (8.253)$$

where ϵ_1 and a_1 are constants. For the validity of the condition (8.251) we require that

$$\epsilon_1(a/a_1)^{-3} \ll 1. \quad (8.254)$$

Hence the solutions approach the extremum at $X = X_0$ with $F_{,X}(X_0) = 0$. The field energy density is given by

$$\rho_\phi \simeq -F_0 + 4F_2X_0^2\epsilon_1(a/a_1)^{-3}, \quad (8.255)$$

where we have neglected the term higher than $\epsilon_1(a/a_1)^{-3}$. For the positivity of the energy density we require that $F_0 < 0$. The equation of state of k-essence is

$$w_\phi \equiv \frac{P_\phi}{\rho_\phi} = - \left[1 + \frac{4F_2}{(-F_0)} X_0^2 \epsilon_1 \left(\frac{a}{a_1} \right)^{-3} \right]^{-1}, \quad (8.256)$$

which approaches the de Sitter value $w_\phi \rightarrow -1$ at late times [$\epsilon_1 (a/a_1)^{-3} \rightarrow 0$]. Recall that the above results are valid under the condition (8.254). It is possible to realize $w_\phi \simeq 0$ during the matter era provided that the condition $4F_2X_0^2/(-F_0) \gg 1$ is satisfied.

The quantities ξ_1 and ξ_2 defined in Eq. (8.30) are given by $\xi_1 = 2F_2(3X - X_0)$ and $\xi_2 = 2F_2(X - X_0)$, respectively. The condition (8.34) to avoid the superluminal sound speed gives $F_2 \geq 0$. Hence the stability conditions (8.30) for quantum fluctuations are satisfied for $X \geq X_0$. Since $\epsilon_1 > 0$ in this case, the field energy density (8.255) remains to be positive. The sound speed squared is given by

$$c_s^2 = \frac{X - X_0}{3X - X_0} = \frac{1}{2}\epsilon_1 \left(\frac{a}{a_1} \right)^{-3}, \quad (8.257)$$

which is much smaller than unity under the condition (8.254). In the asymptotic future it decreases toward $c_s^2 = 0$.

This property is different from the GCG model in which c_s^2 is small at early times and becomes larger at late times. Recall that this growth of c_s^2 is the reason why the GCG model is not compatible with the observations of large-scale structure, while the above single-field k-essence model avoids this problem. The generalization of the model (8.250) satisfying the condition $c_s^2 \ll 1$ has been studied in Ref. [342]. There is another interesting unified model of dark energy and dark matter using the Bose–Einstein condensation, see Ref. [343] for details.

8.7 Future singularities

In the context of General Relativity, if the equation of state of dark energy is less than -1 , the Universe reaches a big-rip singularity with a finite time. This case corresponds to the violation of the null energy condition $\rho + P \geq 0$. The big-rip singularity is called the type I singularity in Ref. [344], which has the following behavior:

$$(I) \text{ Type I: } a \rightarrow \infty, \rho \rightarrow \infty, |P| \rightarrow \infty, \text{ as } t \rightarrow t_s. \quad (8.258)$$

Different types of future singularities appear [345, 346, 347, 348, 349] at a finite time even when the null energy condition is not violated. For example, there is a sudden future singularity [345] characterized by

$$(II) \text{ Type II: } a \rightarrow a_s, \rho \rightarrow \rho_s, |P| \rightarrow \infty, \text{ as } t \rightarrow t_s. \quad (8.259)$$

See Ref. [350] for an early related work. There are also other types of future singularities such as

$$(III) \text{ Type III : } a \rightarrow a_s, \rho \rightarrow \infty, |P| \rightarrow \infty, \quad \text{as } t \rightarrow t_s, \quad (8.260)$$

$$(IV) \text{ Type IV : } a \rightarrow a_s, \rho \rightarrow 0, |P| \rightarrow 0,$$

$$\text{higher derivatives of } H \text{ diverge,} \quad \text{as } t \rightarrow t_s. \quad (8.261)$$

In the following we shall discuss the cases in which such singularities arise.

Let us consider a fluid where the pressure P and the density ρ have the following relation

$$P = -\rho - f(\rho), \quad (8.262)$$

where $f(\rho)$ is a function with respect to ρ . A function of the form $f(\rho) \propto \rho^\alpha$ was first considered in Ref. [351] in the context of inflationary cosmology. Substituting Eq. (8.262) into the continuity equation (2.20), we find that the scale factor is given by

$$a = a_0 \exp \left(\frac{1}{3} \int \frac{d\rho}{f(\rho)} \right), \quad (8.263)$$

where a_0 is a constant. If a fluid with the equation of state (8.262) is the dominant component in the flat FLRW Universe then one has $3H^2 = \kappa^2 \rho$, so that the cosmic time t can be expressed as

$$t = \int \frac{d\rho}{\kappa \sqrt{3\rho} f(\rho)}. \quad (8.264)$$

In order to show an example of the type II singularity, we take the following function

$$f(\rho) = A(\rho_0 - \rho)^{-\gamma}, \quad (8.265)$$

where A , ρ_0 , and $\gamma (> 0)$ are constants. In the limit that $\rho \rightarrow \rho_0$, we have $|P| \rightarrow \infty$ because of the divergence of $f(\rho)$. From Eq. (8.263) the scale factor evolves as

$$a = a_0 \exp \left[-\frac{(\rho_0 - \rho)^{\gamma+1}}{3A(\gamma+1)} \right], \quad (8.266)$$

which means that $a \rightarrow a_0$ for $\rho \rightarrow \rho_0$. The Hubble parameter $H = \dot{a}/a \propto \sqrt{\rho}$ is finite, so that $\dot{a}$ also remains finite. Meanwhile the second derivative $\ddot{a}$ is divergent as $\rho \rightarrow \rho_0$ because of the divergence of the pressure P [see Eq. (2.19)].

From Eq. (8.264) we obtain the following relation around $\rho \sim \rho_0$:

$$t \simeq t_s - \frac{(\rho_0 - \rho)^{\gamma+1}}{\kappa A \sqrt{3\rho_0}(\gamma+1)}, \quad (8.267)$$

where t_s is an integration constant. We then have $t \rightarrow t_s$ in the limit $\rho \rightarrow \rho_0$.

From the above discussion the function $f(\rho)$ in Eq. (8.265) gives rise to the type II singularity. The equation of state is given by $w = P/\rho = -1 - A/[\rho(\rho_0 - \rho)^\gamma]$, so that $w > -1$ for $A < 0$ and $(0 <) \rho < \rho_0$. This means that the sudden singularity is present even for a non-phantom dark energy ($w > -1$).

In order to discuss the type III singularity we take the following function

$$f(\rho) = B\rho^\alpha, \quad (8.268)$$

where B and $\alpha (> 0)$ are constants. From Eq. (8.263) it follows that

$$a = a_0 \exp \left[\frac{\rho^{1-\alpha}}{3(1-\alpha)B} \right], \quad (8.269)$$

and

$$t = t_s + \frac{2}{\sqrt{3\kappa}B} \frac{\rho^{-\alpha+1/2}}{1-2\alpha} \quad (\alpha \neq 1/2), \quad (8.270)$$

$$t = t_s + \frac{\ln \rho}{\sqrt{3\kappa}B} \quad (\alpha = 1/2), \quad (8.271)$$

where t_s is an integration constant.

When $\alpha > 1$ the scale factor is finite even for $\rho \rightarrow \infty$. Meanwhile, when $\alpha < 1$, one has $a \rightarrow \infty$ ($a \rightarrow 0$) as $\rho \rightarrow \infty$ for $B > 0$ ($B < 0$). If $\alpha > 1/2$ the divergence of ρ occurs at the time t_s , whereas, if $\alpha \leq 1/2$, ρ diverges in the infinite future or past. Since the equation of state of dark energy is given by $w = -1 - B\rho^{\alpha-1}$, it follows that $w > -1$ ($w < -1$) for $B < 0$ ($B > 0$). From the above discussion, we can classify the singularities as follows.

1. $\alpha > 1$
There is a type III singularity with $w > -1$ for $B < 0$.
2. $1/2 < \alpha < 1$
There is a type I future singularity with $w < -1$ for $B > 0$. When $B < 0$, one has $a \rightarrow 0$ as $\rho \rightarrow \infty$. Hence if the singularity exists in the future (past), we may call it the Big Crunch (Big Bang) singularity.
3. $0 < \alpha \leq 1/2$
There is no finite future singularity.

Similarly it was shown in Ref. [344] that the type IV singularity is present for the function $f(\rho) = AB\rho^{\alpha+\beta}/(A\rho^\alpha + B\rho^\beta)$.

Finally we note that the Ricci scalar $R = \kappa^2(\rho - 3P)$ diverges as $t \rightarrow t_s$ for the type I, II, III singularities. In such cases higher-order curvature terms are likely to be important as R blows up. In Ref. [344] it was found that quantum corrections coming from conformal anomaly can moderate the finite-time singularities.

It was also shown in Refs. [352, 353] that non-perturbative quantum geometric effects appearing in loop quantum cosmology [354] can lead to the resolution of singularities due to the ρ^2 modification of the effective Friedmann equation [355].

8.8 Problems

- 8.1** The k-essence model with the Lagrangian density (8.27) possesses scaling solutions. Show that the k-essence model with $P(\phi, X) = K(\phi)p(X)$ and $K(\phi) = V_0\phi^{-2}$ has a scaling solution.
- 8.2** Derive Eqs. (8.53)–(8.55).
- 8.3** Perturb Eqs. (8.82)–(8.84) about the fixed points (e), (a), (c) in Table 8.1 and obtain the eigenvalue of the 3×3 Jacobian matrix for each point.
- 8.4** For the Lagrangian density (8.195) let us consider linear perturbations δx , δy , and δY about fixed points (x_c, y_c) , i.e. $x = x_c + \delta x$, $y = y_c + \delta y$, and $\delta Y = Y_c + \delta Y$ in the absence of radiation ($u = 0$). Show that the eigenvalues for the Jacobian matrix $\mathcal{M}$ of perturbations for the fixed points A and B are given by Eqs. (8.213) and (8.219), respectively.

Dark energy as a modification of gravity

In this chapter, we discuss “modified gravity” models in which the origin of dark energy is identified as a modification of gravity. This includes $f(R)$ gravity, scalar-tensor theories, Gauss–Bonnet gravity, and braneworld models of dark energy. In these theories one modifies the laws of gravity so that the late-time accelerated expansion of the Universe is realized without recourse to an explicit dark energy matter component.

Clearly, the modification to the laws of gravity is in general severely restricted from local gravity constraints and from observational constraints. It is of interest to understand how much deviation from the Λ CDM model can be allowed in such modified gravity models.

9.1 $f(R)$ gravity

One of the simplest modified gravity models is the so-called $f(R)$ gravity in which the 4-dimensional action is given by some general function $f(R)$ of the Ricci scalar R :

$$S = \frac{1}{2\kappa^2} \int d^4x \sqrt{-g} f(R) + S_m(g_{\mu\nu}, \Psi_m), \quad (9.1)$$

where as usual $\kappa^2 = 8\pi G$, and S_m is a matter action with matter fields Ψ_m . Here G is a *bare* gravitational constant: we will see that the observed value will in general be different. The matter fields in S_m obey standard conservation equations and therefore the metric $g_{\mu\nu}$ corresponds to the physical frame (which here is the Jordan frame). There are two approaches to derive field equations from the action (9.1).

- **(I) The metric formalism**

The first approach is the so-called metric formalism in which the connections $\Gamma_{\beta\gamma}^\alpha$ are the usual connections defined in terms of the metric $g_{\mu\nu}$. The field equations can be obtained

by varying the action (9.1) with respect to $g_{\mu\nu}$:

$$F(R)R_{\mu\nu}(g) - \frac{1}{2}f(R)g_{\mu\nu} - \nabla_\mu \nabla_\nu F(R) + g_{\mu\nu}\square F(R) = \kappa^2 T_{\mu\nu}, \quad (9.2)$$

where $F(R) \equiv \partial f / \partial R$ (we also use the notation $f_{,R} \equiv \partial f / \partial R$, $f_{,RR} \equiv \partial^2 f / \partial R^2$), and $T_{\mu\nu}$ is the matter energy-momentum tensor. The steps to derive this equation are similar to those employed to obtain the field equation (6.7). The trace of Eq. (9.2) is given by

$$3\square F(R) + F(R)R - 2f(R) = \kappa^2 T, \quad (9.3)$$

where $T = g^{\mu\nu}T_{\mu\nu} = -\rho + 3P$. Here ρ and P are the energy density and the pressure of matter, respectively.

• **(II) The Palatini formalism**

The second approach is the so-called Palatini¹ formalism in which $\Gamma_{\beta\gamma}^\alpha$ and $g_{\mu\nu}$ are treated as independent variables. Varying the action (9.1) with respect to $g_{\mu\nu}$ gives

$$F(R)R_{\mu\nu}(\Gamma) - \frac{1}{2}f(R)g_{\mu\nu} = \kappa^2 T_{\mu\nu}, \quad (9.4)$$

where $R_{\mu\nu}(\Gamma)$ is the Ricci tensor corresponding to the connections $\Gamma_{\beta\gamma}^\alpha$. In general this is different from the Ricci tensor $R_{\mu\nu}(g)$ corresponding to the metric connections. Taking the trace of Eq. (9.4), we obtain

$$F(R)R - 2f(R) = \kappa^2 T, \quad (9.5)$$

where $R(T) = g^{\mu\nu}R_{\mu\nu}(\Gamma)$ is directly related to T . Taking the variation of the action (9.1) with respect to the connection, and using Eq. (9.4), we find

$$\begin{aligned} R_{\mu\nu}(g) - \frac{1}{2}g_{\mu\nu}R(g) &= \frac{\kappa^2 T_{\mu\nu}}{F} - \frac{FR(T) - f}{2F}g_{\mu\nu} + \frac{1}{F}(\nabla_\mu \nabla_\nu F - g_{\mu\nu}\square F) \\ &\quad - \frac{3}{2F^2} \left[\partial_\mu F \partial_\nu F - \frac{1}{2}g_{\mu\nu}(\nabla F)^2 \right]. \end{aligned} \quad (9.6)$$

In General Relativity we have $f(R) = R - 2\Lambda$ and $F(R) = 1$, so that the term $\square F(R)$ in Eq. (9.3) vanishes. In this case both the metric and the Palatini formalisms give the relation $R = -\kappa^2 T = \kappa^2(\rho - 3P)$, which means that the Ricci scalar R is directly determined by the matter (the trace T).

In modified gravity models where $F(R)$ is a function of R , the term $\square F(R)$ does not vanish in Eq. (9.3). This means that, in the metric formalism, there is a propagating scalar degree of freedom, $\psi \equiv F(R)$. The trace equation (9.3) governs the dynamics of the scalar field ψ (dubbed “scalaron” [8]). In the Palatini formalism the kinetic term $\square F(R)$ is not present in Eq. (9.5), which means that the scalar-field degree of freedom does not propagate freely.

¹ Attilio Palatini (1889–1949), Italian mathematician, generalized the concept of the variational principle in relativity.

The de Sitter point corresponds to a vacuum solution at which the Ricci scalar is constant. Since $\square F(R) = 0$ at this point, we get

$$F(R)R - 2f(R) = 0, \quad (9.7)$$

which holds for both the metric and the Palatini formalisms. Since the model $f(R) = \alpha R^2$ satisfies this condition, it possesses an exact de Sitter solution [8].

Since the dynamics of $f(R)$ dark energy models is different depending on the two formalisms, we shall discuss the metric and Palatini formalisms separately.

9.1.1 $f(R)$ gravity in the metric formalism

We first discuss the $f(R)$ dark energy models based on the metric formalism.

Already in the early 1980s it was known that the model $f(R) = R + \alpha R^2$ can be responsible for inflation in the early Universe [8]. This comes from the fact that the presence of the quadratic term αR^2 gives rise to an asymptotically exact de Sitter solution. Inflation ends when the term αR^2 becomes smaller than the linear term R . Since the term αR^2 is negligibly small relative to R at the present epoch, this model is not suitable to realize the present cosmic acceleration.

Since a late-time acceleration requires modification for small R , models of the type $f(R) = R - \alpha/R^n$ ($\alpha > 0, n > 0$) were proposed as a candidate for dark energy [53, 54, 55, 356]. While the late-time cosmic acceleration is possible in these models, it has become clear that they do not satisfy local gravity constraints because of the instability associated with negative values of $f_{,RR}$ [357, 358, 359, 360, 361]. Moreover a standard matter epoch is not present because of a large coupling between the Ricci scalar and the non-relativistic matter [272].

Then, what are the conditions for the viability of $f(R)$ dark energy models in the metric formalism? In the following we first present such conditions and then explain why they are required step by step.

- (i) $f_{,R} > 0$ for $R \geq R_0$ (> 0), where R_0 is the Ricci scalar at the present epoch. Strictly speaking, if the final attractor is a de Sitter point with the Ricci scalar R_1 (> 0), then the condition $f_{,R} > 0$ needs to hold for $R \geq R_1$.
This is required to avoid anti-gravity (see later on).
- (ii) $f_{,RR} > 0$ for $R \geq R_0$.
This is required for consistency with local gravity tests [358, 360, 361, 362], for the presence of the matter-dominated epoch [272, 363], and for the stability of cosmological perturbations [364, 365, 366, 367].

- (iii) $f(R) \rightarrow R - 2\Lambda$ for $R \gg R_0$.

This is required for consistency with local gravity tests [267, 368, 369, 370, 371] and for the presence of the matter-dominated epoch [363].

- (iv) $0 < \frac{Rf_{,RR}}{f_{,R}}(r = -2) < 1$ at $r = -\frac{Rf_{,R}}{f} = -2$.

This is required for the stability of the late-time de Sitter point [372, 373, 363].

For example, the model $f(R) = R - \alpha/R^n$ ($\alpha > 0$, $n > 0$) does not satisfy the condition (ii).

Below we list some viable $f(R)$ models that satisfy the above conditions.

$$(A) \quad f(R) = R - \mu R_c (R/R_c)^p \quad \text{with } 0 < p < 1, \quad \mu, R_c > 0, \quad (9.8)$$

$$(B) \quad f(R) = R - \mu R_c \frac{(R/R_c)^{2n}}{(R/R_c)^{2n} + 1} \quad \text{with } n, \mu, R_c > 0, \quad (9.9)$$

$$(C) \quad f(R) = R - \mu R_c \left[1 - (1 + R^2/R_c^2)^{-n} \right] \quad \text{with } n, \mu, R_c > 0, \quad (9.10)$$

$$(D) \quad f(R) = R - \mu R_c \tanh(R/R_c) \quad \text{with } \mu, R_c > 0. \quad (9.11)$$

The models (A), (B), (C), and (D) have been proposed in Refs. [363], [368], [369], and [371], respectively. A model similar to (D) has also been proposed in Ref. [370], while a generalized model encompassing (B) and (C) has been studied in Ref. [375]. In the model (A), the power p needs to be close to 0 to satisfy the condition (iii). In the models (B) and (C) the function $f(R)$ asymptotically behaves as $f(R) \rightarrow R - \mu R_c [1 - (R^2/R_c^2)^{-n}]$ for $R \gg R_c$ and hence the condition (iii) can be satisfied even for $n = \mathcal{O}(1)$. In the model (D) the function $f(R)$ rapidly approaches $f(R) \rightarrow R - \mu R_c$ in the region $R \gg R_c$. These models satisfy $f(R=0) = 0$, so the cosmological constant vanishes in the flat space-time.

9.1.2 Cosmological dynamics of $f(R)$ dark energy models in the metric formalism

Let us consider cosmological dynamics of $f(R)$ gravity in the metric formalism. It is possible to carry out a general analysis without specifying the form of $f(R)$. In the flat FLRW spacetime the Ricci scalar is given by

$$R = 6(2H^2 + \dot{H}), \quad (9.12)$$

where H is as usual the Hubble parameter. As a matter action S_m we take into account non-relativistic matter and radiation, which satisfy the usual conservation equations $\dot{\rho}_m + 3H\rho_m = 0$ and $\dot{\rho}_r + 4H\rho_r = 0$ respectively. From Eqs. (9.2) and

(9.3) we obtain the following equations

$$3FH^2 = \kappa^2(\rho_m + \rho_r) + (FR - f)/2 - 3H\dot{F}, \quad (9.13)$$

$$-2F\dot{H} = \kappa^2[\rho_m + (4/3)\rho_r] + \ddot{F} - H\dot{F}. \quad (9.14)$$

We introduce the dimensionless variables:

$$x_1 \equiv -\frac{\dot{F}}{HF}, \quad x_2 \equiv -\frac{f}{6FH^2}, \quad x_3 \equiv \frac{R}{6H^2}, \quad x_4 \equiv \frac{\kappa^2\rho_r}{3FH^2}, \quad (9.15)$$

together with the following quantities

$$\Omega_m \equiv \frac{\kappa^2\rho_m}{3FH^2} = 1 - x_1 - x_2 - x_3 - x_4, \quad \Omega_r \equiv x_4, \quad \Omega_{DE} \equiv x_1 + x_2 + x_3. \quad (9.16)$$

It is straightforward to derive the following differential equations [363]:

$$\frac{dx_1}{dN} = -1 - x_3 - 3x_2 + x_1^2 - x_1x_3 + x_4, \quad (9.17)$$

$$\frac{dx_2}{dN} = \frac{x_1x_3}{m} - x_2(2x_3 - 4 - x_1), \quad (9.18)$$

$$\frac{dx_3}{dN} = -\frac{x_1x_3}{m} - 2x_3(x_3 - 2), \quad (9.19)$$

$$\frac{dx_4}{dN} = -2x_3x_4 + x_1x_4, \quad (9.20)$$

where $N = \ln a$ and

$$m \equiv \frac{d \ln F}{d \ln R} = \frac{Rf_{,RR}}{f_{,R}}, \quad (9.21)$$

$$r \equiv -\frac{d \ln f}{d \ln R} = -\frac{Rf_{,R}}{f} = \frac{x_3}{x_2}. \quad (9.22)$$

From Eq. (9.22) one can express R as a function of x_3/x_2 . Since m is a function of R , it follows that m is a function of r , i.e. $m = m(r)$. The Λ CDM model, $f(R) = R - 2\Lambda$, corresponds to $m = 0$. Hence the quantity m characterizes the deviation from the Λ CDM model. Note also that the model, $f(R) = \alpha R^{1+m} - 2\Lambda$, gives a constant value of m . The analysis using Eqs. (9.17)–(9.20) is sufficiently general in the sense that the form of $f(R)$ does not need to be specified.

The effective equation of state of the system defined in Eq. (4.96) is

$$w_{\text{eff}} = -\frac{1}{3}(2x_3 - 1). \quad (9.23)$$

In the absence of radiation ($x_4 = 0$) the fixed points for the dynamical system (9.17)–(9.20) are

$$P_1 : (x_1, x_2, x_3) = (0, -1, 2), \quad \Omega_m = 0, \quad w_{\text{eff}} = -1, \quad (9.24)$$

$$P_2 : (x_1, x_2, x_3) = (-1, 0, 0), \quad \Omega_m = 2, \quad w_{\text{eff}} = 1/3, \quad (9.25)$$

$$P_3 : (x_1, x_2, x_3) = (1, 0, 0), \quad \Omega_m = 0, \quad w_{\text{eff}} = 1/3, \quad (9.26)$$

$$P_4 : (x_1, x_2, x_3) = (-4, 5, 0), \quad \Omega_m = 0, \quad w_{\text{eff}} = 1/3, \quad (9.27)$$

$$P_5 : (x_1, x_2, x_3) = \left(\frac{3m}{1+m}, -\frac{1+4m}{2(1+m)^2}, \frac{1+4m}{2(1+m)} \right),$$

$$\Omega_m = 1 - \frac{m(7+10m)}{2(1+m)^2}, \quad w_{\text{eff}} = -\frac{m}{1+m}, \quad (9.28)$$

$$P_6 : (x_1, x_2, x_3) = \left(\frac{2(1-m)}{1+2m}, \frac{1-4m}{m(1+2m)}, -\frac{(1-4m)(1+m)}{m(1+2m)} \right),$$

$$\Omega_m = 0, \quad w_{\text{eff}} = \frac{2-5m-6m^2}{3m(1+2m)}. \quad (9.29)$$

The points P_5 and P_6 lie on the line $m(r) = -r - 1$ in the (r, m) plane.

It is important to remark that $P_{5,6}$ represent actually two *families* of points. In fact $m(x_3/x_2)$ is a function of the coordinates and for each model one has to solve the three equations

$$\{x_1, x_2, x_3\} = \left\{ \frac{3m(x_3/x_2)}{1+m(x_3/x_2)}, -\frac{1+4m(x_3/x_2)}{2[1+m(x_3/x_2)]^2}, \frac{1+4m(x_3/x_2)}{2[1+m(x_3/x_2)]} \right\}, \quad (9.30)$$

for P_5 and an analogous set for P_6 . We will call $m_{5,6}$ the solutions of these equations and

$$m'_{5,6} \equiv \left. \frac{dm}{dr} \right|_{m_{5,6}} \quad (9.31)$$

their derivatives. For a given model there are several fixed points of type $P_{5,6}$, all of them lying on the critical line $m = -r - 1$. For simplicity, however, we will refer to points $P_{5,6}$ in the following discussion as if they were single points because for every viable cosmological model only one point of each family really matters.

Among the six fixed points we have presented above, only the point P_5 can be used for the matter-dominated epoch. Since in this case we require $\Omega_m \simeq 1$ and $w_{\text{eff}} \simeq 0$, this implies that m is close to 0. In the (r, m) plane this point exists around $(r, m) = (-1, 0)$. The point P_2 corresponds to the ϕ MDE [17], but in this case the ϕ MDE cannot be responsible for the matter fixed point because $w_{\text{eff}}(P_2) = 1/3$. In $f(R)$ gravity, a scalar field degree of freedom has a large coupling ($Q = -1/\sqrt{6}$) with non-relativistic matter in the Einstein frame (as we will see later).

One can study the stability of the above fixed points by considering perturbations δx_i ($i = 1, 2, 3$) around them [363]. For the point P_5 the eigenvalues for the 3×3 Jacobian matrix of perturbations are

$$3(1 + m'_5), \quad \frac{-3m_5 \pm \sqrt{m_5(256m_5^3 + 160m_5^2 - 31m_5 - 16)}}{4m_5(m_5 + 1)}, \quad (9.32)$$

where $r_5 \approx -1$. In the limit $|m_5| \ll 1$ the latter two eigenvalues reduce to $-3/4 \pm \sqrt{-1/m_5}$. The $f(R)$ models with $m_5 < 0$ exhibit a divergence of the eigenvalues as $m_5 \rightarrow -0$, in which case the system cannot remain for a long time around the point P_5 . For example the model $f(R) = R - \alpha/R^n$ with $n > 0$ and $\alpha > 0$ falls into this category. On the other hand, if $0 < m_5 < 0.327$, the latter two eigenvalues in Eq. (9.32) are complex with negative real parts. Then, provided that $m'_5 > -1$, the point P_5 corresponds to a saddle point with a damped oscillation. Hence the Universe can evolve toward the point P_5 from the radiation era and then can leave for the late-time acceleration. In summary the condition for the existence of the saddle matter era is

$$m(r) \approx +0, \quad \frac{dm}{dr} > -1, \quad \text{at } r = -1. \quad (9.33)$$

The first condition implies that the $f(R)$ models need to be close to the Λ CDM model during the matter-dominated epoch.

The points P_1 and P_6 can lead to the late-time cosmic acceleration. The point P_1 corresponds to a de Sitter solution at which $r = -2$. In fact the condition (9.7) is satisfied in this case. The eigenvalues for the 3×3 matrix of perturbations about the point P_1 are

$$-3, \quad -\frac{3}{2} \pm \frac{\sqrt{25 - 16/m_1}}{2}, \quad (9.34)$$

where $m_1 = m(r = -2)$. This shows that the condition for the stability of the de Sitter point P_1 is given by

$$0 < m(r = -2) < 1, \quad (9.35)$$

which corresponds to the condition (iv) given in Section 9.1.1. The trajectories which start from the saddle matter point P_5 [satisfying the condition (9.33)] and then approach the stable de Sitter point P_1 [satisfying the condition (9.35)] are cosmologically viable.

The point P_6 is on the line $m(r) = -r - 1$. It can satisfy the condition for the cosmic acceleration ($w_{\text{eff}} < -1/3$) provided that $m_6 < -(1 + \sqrt{3})/2$, or $-1/2 <$

$m_6 < 0$, or $m_6 > (\sqrt{3} - 1)/2$. The eigenvalues for perturbations are given by

$$-4 + \frac{1}{m_6}, \quad \frac{2 - 3m_6 - 8m_6^2}{m_6(1 + 2m_6)}, \quad -\frac{2(m_6^2 - 1)(1 + m_6')}{m_6(1 + 2m_6)}. \quad (9.36)$$

We then find that P_6 is stable and accelerated in the following four regions:

[I] $m_6' > -1$

- (a) $m_6 < -(1 + \sqrt{3})/2$: P_6 is accelerated with the effective equation of state, $w_{\text{eff}} > -1$. One has $w_{\text{eff}} \rightarrow -1$ in the limit $m_6 \rightarrow -\infty$.
- (b) $-1/2 < m_6 < 0$: P_6 corresponds to a strongly phantom behavior with $w_{\text{eff}} < -7.6$.
- (c) $m_6 \geq 1$: P_6 corresponds to a slightly phantom behavior with $-1.07 < w_{\text{eff}} \leq -1$. One has $w_{\text{eff}} \rightarrow -1$ in the limit $m_6 \rightarrow +\infty$ and $m_6 \rightarrow 1$.

[II] $m_6' < -1$

When $m_6' < -1$, the point P_6 is stable and accelerated in the following region.

- (d) $(\sqrt{3} - 1)/2 < m_6 < 1$: P_6 corresponds to a non-phantom behavior with $w_{\text{eff}} > -1$.

Recall that the matter point P_5 needs to satisfy the condition $m_5 \approx +0$ and $m_5'(r) > -1$ at $r = -1$ and that both P_5 and P_6 are on the line $m = -r - 1$. If we consider curves connecting P_5 to P_6 , it is not possible to realize the trajectories to the point P_6 in the regions (a), (b), (c) satisfying the condition $m_6'(r) > -1$. In other words, once a trajectory crosses the line $m = -r - 1$ with the tangent $m_5'(r) > -1$, then it crosses the same line again with the tangent $m_6'(r) < -1$. See the curve (iv) in Fig. 9.1 for illustration. From the above argument the viable trajectories evolve from the point P_5 to the point P_6 in the region (d).

In summary we have only two qualitatively different viable cases:

- Class A: Models that link P_5 with P_1 ($r = -2, 0 < m < 1$).
- Class B: Models that link P_5 with P_6 ($m = -r - 1, (\sqrt{3} - 1)/2 < m < 1$).

Let us consider a couple of viable $f(R)$ models in the (r, m) plane. The Λ CDM model, $f(R) = R - 2\Lambda$, corresponds to $m = 0$, in which case the trajectory is a straight line from P_5 : $(r, m) = (-1, 0)$ to P_1 : $(r, m) = (-2, 0)$. The trajectory (ii) in Fig. 9.1 represents the following model [267]

$$f(R) = (R^b - \Lambda)^c, \quad (9.37)$$

which corresponds to the straight line $m(r) = [(1 - c)/c]r + b - 1$ in the (r, m) plane. The existence of a saddle matter epoch requires the condition $c \geq 1$ and $bc \approx 1$. The trajectory (iii) represents the model $f(R) = R - \mu R_c (R/R_c)^p$ ($0 < p < 1, \mu, R_c > 0$) [363, 376], which corresponds to the curve $m = p(1 + r)/r$. These models fall into Class A.

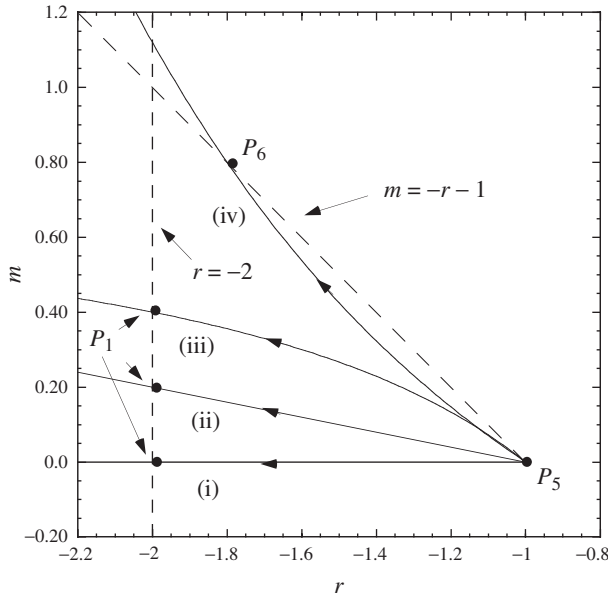


Figure 9.1 Four trajectories in the (r, m) plane. Each trajectory corresponds to the models: (i) Λ CDM, (ii) $f(R) = (R^b - \Lambda)^c$, (iii) $f(R) = R - \mu R_c(R/R_c)^p$ ($0 < p < 1, \mu, R_c > 0$), and (iv) $m(r) = -C(r+1)(r^2 + ar + b)$.

The models (9.9) and (9.10) have the same asymptotic form $f(R) \simeq R - \mu R_c[1 - (R/R_c)^{-2n}]$ in the region $R \gg R_c$. In this region these models behave as

$$m(r) = C(-r-1)^{2n+1}, \quad (9.38)$$

where $C = 2n(2n+1)/\mu^{2n}$. The parameter $m(r)$ rapidly approaches 0 in the limit $r \rightarrow -1$ because of the presence of the power $2n+1$ larger than 1. As we will see later, this small value of m in the region of high density is required for consistency with local gravity constraints. These models can be categorized by either Class A or Class B. The trajectory (iv) in Fig. 9.1 shows the model $m(r) = -C(r+1)(r^2 + ar + b)$, which belongs to Class B. We require the conditions $m'(-1) = -C(1-a+b) > -1$ and $m'(-2) = C(3a-b-8) < -1$ for the transition from the matter point P_5 to the stable accelerated point P_6 . The models shown in Fig. 9.1 are a couple of representative models giving viable cosmological evolution.

In the presence of the radiation ($x_4 \neq 0$) we have the following two points in addition to the points P_{1-6} :

- $P_7 : (x_1, x_2, x_3, x_4) = (0, 0, 0, 1), \Omega_m = 0, \Omega_{DE} = 0, w_{\text{eff}} = 1/3, \quad (9.39)$

- $P_8 : (x_1, x_2, x_3, x_4) = \left(\frac{4m}{1+m}, -\frac{2m}{(1+m)^2}, \frac{2m}{1+m}, \frac{1-2m-5m^2}{(1+m)^2} \right),$

$$\Omega_m = 0, \quad \Omega_{DE} = \frac{2m(2+3m)}{(1+m)^2}, \quad w_{\text{eff}} = \frac{1-3m}{3+3m}. \quad (9.40)$$

Here P_7 is a standard radiation point. For constant m the eigenvalues of P_7 are given by 4, 4, 1, -1 , which means that P_7 is a saddle in this case. The point P_8 corresponds to a new radiation era having non-zero dark energy. Since the dark energy density is constrained to be $\Omega_{\text{DE}} < 0.045$ from the BBN bound, P_8 is acceptable as a radiation point only for $m_8 \lesssim 0.01$.

The eigenvalues of P_8 are given by

$$1, \quad 4(1 + m'_8), \quad \frac{m_8 - 1 \pm \sqrt{81m_8^2 + 30m_8 - 15}}{2(m_8 + 1)}. \quad (9.41)$$

In the limit $m_8 \rightarrow 0$ the last two values are complex with negative real parts, which shows that P_8 is a saddle point. Then the solutions eventually repel away from the point P_8 , which are followed by one of the fixed points given above. Unlike the matter point P_5 , there are no singularities for the eigenvalues of P_8 in the limit $m_8 \rightarrow 0$. We also note that P_8 is on the line $m = -r - 1$. If the condition for the existence of the matter point P_5 is satisfied (i.e. $m \approx +0$ and $r \approx -1$), there exists a radiation point P_8 in the same region. Then a viable cosmological trajectory starts around the point P_8 with $m \approx +0$ and then connects to the matter point P_5 . Finally the solutions approach either the accelerated point P_1 or P_6 .

The requirement $m \rightarrow +0$ during the radiation and matter eras means that the models need to be close to the Λ CDM model, $f(R) = R - 2\Lambda$, in the region $R \gg R_0$ (where R_0 is the present cosmological Ricci scalar). This corresponds to the condition (iii) listed in Section 9.1.1. Note also that the Ricci scalar $R = 6(2H^2 + \dot{H})$ remains positive from the radiation era to the present epoch, as long as it does not oscillate as in the $f(R) = R + \alpha R^2$ model. Under the condition $f_{,R} > 0$, the requirement $m > 0$ translates into the condition $f_{,RR} > 0$. This is the condition (ii) listed in Section 9.1.1, which is also required for the consistency with local gravity constraints (as we will see later).

For the model (9.9) let us consider the case in which the solutions finally approach the de Sitter point P_1 with the Ricci scalar R_1 . The de Sitter point at $r = -Rf_{,R}/f = -2$ is determined by the value μ :

$$\mu = \frac{(1 + x_1^{2n})^2}{x_1^{2n-1}(2 + 2x_1^{2n} - 2n)}, \quad (9.42)$$

where $x_1 \equiv R_1/R_c$. From the stability condition $0 < m(r = -2) < 1$ we obtain

$$2x_1^{4n} - (2n - 1)(2n + 4)x_1^{2n} + (2n - 1)(2n - 2) > 0. \quad (9.43)$$

When $n = 1$, for example, we have $x_1 > \sqrt{3}$ and $\mu > 8\sqrt{3}/9$. Under Eq. (9.43) one can show that the conditions $f_{,R} > 0$ and $f_{,RR} > 0$ are also satisfied for $R \geq R_1$. For μ and n of the order of unity we find from Eq. (9.42) that R_1 is the same order as R_c . Hence R_c is roughly the same order as the present cosmological Ricci

scalar R_0 . In the region $R \gg R_c$ the model (9.9) is close to the Λ CDM model with the asymptotic form $f(R) \simeq R - \mu R_c [1 - (R/R_c)^{-2n}]$. The deviation from the Λ CDM model becomes important when R decreases to the order of R_c . Note that the model (9.10) also has a similar property.

In order to derive the equation of state of dark energy to confront with SN Ia observations, we rewrite Eqs. (9.13) and (9.14) as follows:

$$3AH^2 = \kappa^2 (\rho_m + \rho_r + \rho_{\text{DE}}) , \quad (9.44)$$

$$-2A\dot{H} = \kappa^2 [\rho_m + (4/3)\rho_r + \rho_{\text{DE}} + P_{\text{DE}}] , \quad (9.45)$$

where A is some constant and

$$\kappa^2 \rho_{\text{DE}} \equiv (1/2)(FR - f) - 3H\dot{F} + 3H^2(A - F) , \quad (9.46)$$

$$\kappa^2 P_{\text{DE}} \equiv \ddot{F} + 2H\dot{F} - (1/2)(FR - f) - (3H^2 + 2\dot{H})(A - F) . \quad (9.47)$$

Defining ρ_{DE} and P_{DE} in the above way, one can show that these satisfy the usual conservation equation

$$\dot{\rho}_{\text{DE}} + 3H(\rho_{\text{DE}} + P_{\text{DE}}) = 0 . \quad (9.48)$$

A similar procedure can be carried out for a more general Lagrangian density, $f(R, \phi, X)$ (see problem 9.1). The dark energy equation of state, $w_{\text{DE}} \equiv P_{\text{DE}}/\rho_{\text{DE}}$, is directly related to the one used in the standard analysis of SN Ia observations. From Eqs. (9.44) and (9.45) it is given by

$$w_{\text{DE}} = -\frac{2A\dot{H} + 3AH^2 + \kappa^2 \rho_r/3}{3AH^2 - \kappa^2(\rho_m + \rho_r)} \simeq \frac{w_{\text{eff}}}{1 - \tilde{\Omega}_m} , \quad (9.49)$$

where

$$\tilde{\Omega}_m \equiv \frac{\kappa^2 \rho_m}{3AH^2} = \frac{F}{A} \Omega_m . \quad (9.50)$$

The last approximate equality in Eq. (9.49) is valid in the regime where the radiation energy density ρ_r is negligible relative to the matter density.

The viable $f(R)$ models approach the Λ CDM model in the past, i.e. $F \rightarrow 1$ as $R \rightarrow \infty$. In order to reproduce the standard matter era for $z \gg 1$, we can choose $A = 1$ in Eqs. (9.44) and (9.45). Another possible choice is $A = F_0$, where F_0 is the present value of F . This choice is suitable if the deviation of F_0 from 1 is small (as in the scalar-tensor theory with a massless scalar field [374]). In both cases the equation of state changes from $w_{\text{DE}} < -1$ to $w_{\text{DE}} > -1$ before reaching the de Sitter attractor for viable $f(R)$ models [368, 267, 371]. This is associated with the decrease of the quantity F with time (coming from the condition $F_{,R} > 0$ with $\dot{R} < 0$). Thus viable $f(R)$ dark energy models give rise to a phantom equation of state without violating stability conditions of the system.

For the cosmological viability of $f(R)$ models the constraint on m is not so severe: m can be of the order of 0.1 around the present epoch. Meanwhile the consistency with local gravity experiments places a tighter bound on m in the region of high density ($R \gg R_0$), which corresponds to the value of $m \ll 10^{-9}$ during radiation and deep matter eras [see Eq. (9.86) in the next subsection]. The models (9.9)–(9.11) are carefully constructed to have a suppressed m in the early cosmological epoch, while an appreciable deviation from the Λ CDM model ($m \lesssim \mathcal{O}(0.1)$) can appear around the present. Note that the model (9.8) does not allow this rapid evolution of m .

Integrating Eqs. (9.17)–(9.20) numerically, one finds that for the models in which m quickly decreases toward $+0$ in the past the denominators in Eqs. (9.18) and (9.19) give rise to the rapid oscillations of the variables x_i because the mass of the scalaron becomes very large. If we start integrating the equations from the deep radiation era, we typically encounter this oscillating behavior [371, 377]. This is associated with the fact that the oscillating mode of perturbations tends to dominate over the matter-induced mode of perturbations unless initial conditions are appropriately chosen [369]. In Section 11.6 we shall discuss the evolution of matter density perturbations for cosmologically viable $f(R)$ models. This also provides useful information to distinguish the $f(R)$ model from other dark energy models.

9.1.3 Local gravity constraints on $f(R)$ gravity models in the metric formalism

Gravity is severely constrained by local gravity tests. In this subsection we discuss local gravity constraints on $f(R)$ gravity models in the metric formalism from the violation of the equivalence principle.

Let us consider local fluctuations on a background characterized by a curvature R_0 and a density ρ_0 . We shall expand Eq. (9.3) in powers of fluctuations under the weak field approximation. We decompose the quantities $\psi \equiv F(R)$, $g_{\mu\nu}$, and $T_{\mu\nu}$ into the background part and the perturbed part: $\psi = \psi_0(1 + \delta_\psi)$, $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$, and $T_{\mu\nu} = T_{\mu\nu}^{(0)} + \delta T_{\mu\nu}$, where we have used the approximation $g_{\mu\nu}^{(0)} \approx \eta_{\mu\nu}$. Then the trace equation (9.3) yields [360, 362] (see problem 9.2)

$$\left(\frac{\partial^2}{\partial t^2} - \nabla^2 \right) \delta_\psi + M_\psi^2 \delta_\psi = -\frac{\kappa^2}{3\psi_0} \delta T, \quad (9.51)$$

where $\delta T \equiv \eta^{\mu\nu} \delta T_{\mu\nu}$ and

$$M_\psi^2 \equiv \frac{1}{3} \left[\frac{f_{,R}(R_0)}{f_{,RR}(R_0)} - R_0 \right] = \frac{R_0}{3} \left[\frac{1}{m(R_0)} - 1 \right]. \quad (9.52)$$

In the case of the homogeneous and isotropic cosmological setting (without a Hubble friction) where δ_ψ is a function of the cosmic time t only, Eq. (9.51) reduces to

$$\ddot{\delta}_\psi + M_\psi^2 \delta_\psi = \frac{\kappa^2}{3\psi_0} \rho, \quad (9.53)$$

where $\rho \equiv -\delta T$. For the models where the deviation from the Λ CDM model is small, we have $m(R_0) \ll 1$ so that $|M_\psi^2|$ is much larger than R_0 . If $M_\psi^2 < 0$, the perturbation δ_ψ exhibits a violent instability. Hence the condition $M_\psi^2 \simeq f_{,R}(R_0)/(3f_{,RR}(R_0)) > 0$ is needed for the stability of cosmological perturbations. Since $f_{,R}(R_0) > 0$ to avoid anti-gravity (see below), we require the condition $f_{,RR}(R_0) > 0$ [364, 365, 366, 367].

Let us consider a spherically symmetric body with mass M_c , constant density ρ , radius r_c , and vanishing density outside the body. In this case δ_ψ is a function of the distance r from the center of the body,² so that Eq. (9.51) inside the body yields

$$\frac{d^2}{dr^2} \delta_\psi + \frac{2}{r} \frac{d}{dr} \delta_\psi - M_\psi^2 \delta_\psi = -\frac{\kappa^2}{3\psi_0} \rho. \quad (9.54)$$

Outside the body the r.h.s. of Eq. (9.54) vanishes. Then the solution of the perturbation δ_ψ for $M_\psi^2 > 0$ is given by

$$(\delta_\psi)_{r>r_c} = C_1 \frac{e^{-M_\psi r}}{r} + C_2 \frac{e^{M_\psi r}}{r}, \quad (9.55)$$

$$(\delta_\psi)_{r<r_c} = C_3 \frac{e^{-M_\psi r}}{r} + C_4 \frac{e^{M_\psi r}}{r} + \frac{8\pi G\rho}{3\psi_0 M_\psi^2}, \quad (9.56)$$

where C_1, C_2, C_3, C_4 are integration constants. From the requirement that ψ takes the background value ψ_0 at infinity, we require that $(\delta_\psi)_{r>r_c} \rightarrow 0$ as $r \rightarrow \infty$. This sets C_2 to be 0. From the regularity condition at $r = 0$ we require that $C_4 = -C_3$. We match two solutions (9.55) and (9.56) by using the boundary conditions $(\delta_\psi)_{r>r_c}(r_c) = (\delta_\psi)_{r<r_c}(r_c)$ and $(\delta_\psi)'_{r>r_c}(r_c) = (\delta_\psi)'_{r<r_c}(r_c)$. If $M_\psi r_c \ll 1$ we obtain the following solutions [362]

$$(\delta_\psi)_{r>r_c} \simeq \frac{2GM_c}{3\psi_0 r} e^{-M_\psi r}, \quad (9.57)$$

$$(\delta_\psi)_{r<r_c} \simeq \frac{4\pi G\rho}{3\psi_0} \left(r_c^2 - \frac{r^2}{3} \right). \quad (9.58)$$

The first-order solution for the fluctuation $h_{\mu\nu}$ of the metric $\tilde{g}_{\mu\nu} = \psi_0(\eta_{\mu\nu} + h_{\mu\nu})$ follows from the first-order linearized Einstein equations,

² In this subsection we use r to denote the distance instead of the quantity $(-Rf_{,R}/f)$ defined in Eq. (9.22).

$\delta\tilde{G}_{\mu\nu} = 8\pi(G/\psi_0)\delta\tilde{T}_{\mu\nu}$, in the Einstein gravity. Note that the gravitational constant G is modified because of the transformation of the metric. This gives the standard result, $h_{00} = 2GM_c/(\psi_0 r)$ and $h_{ij} = 2GM_c/(\psi_0 r)\delta_{ij}$. The actual metric $g_{\mu\nu}$ is given by

$$g_{\mu\nu} = \frac{\tilde{g}_{\mu\nu}}{\psi} \simeq \eta_{\mu\nu} + h_{\mu\nu} - \delta_\psi \eta_{\mu\nu}. \quad (9.59)$$

Using the solution (9.57) outside the body, we find that the (00) and (ii) components of the metric $g_{\mu\nu}$ are

$$g_{00} \simeq -1 + \frac{2G_{\text{eff}}^{(N)} M_c}{r}, \quad g_{ii} \simeq 1 + \frac{2G_{\text{eff}}^{(N)} M_c}{r} \gamma, \quad (9.60)$$

where $G_{\text{eff}}^{(N)}$ and γ are the effective gravitational coupling and the post-Newtonian parameter, respectively, defined by

$$G_{\text{eff}}^{(N)} \equiv \frac{G}{\psi_0} \left(1 + \frac{1}{3} e^{-M_\psi r} \right), \quad \gamma \equiv \frac{3 - e^{-M_\psi r}}{3 + e^{-M_\psi r}}. \quad (9.61)$$

Assuming conventionally the bare constant G to be positive, a positive $G_{\text{eff}}^{(N)}$ demands $\psi_0 = F_0 > 0$. In principle one could have F negative in the past and positive today but generally speaking whenever F crosses zero singularities appear [214]. Therefore we must assume $F > 0$ during the whole cosmological evolution. This explains the condition (i) of Section 9.1.1.

In the massless limit $M_\psi \rightarrow 0$ these reduce to $G_{\text{eff}}^{(N)} = (4/3)(G/\psi_0)$ and $\gamma = 1/2$. Since $\psi_0 \delta_\psi = f_{,RR}(R_0) \delta R$, it follows that

$$\delta R = \frac{f_{,R}(R_0)}{f_{,RR}(R_0)} \delta_\psi. \quad (9.62)$$

For the validity of the above linear expansion we require that $\delta R \ll R_0$, which translates into the condition $\delta_\psi \ll m(R_0)$. Using the value $\delta_\psi \simeq 2GM_c/(3\psi_0 r_c)$ at $r = r_c$, this condition is simply expressed as

$$m(R_0) \gg \Phi_c, \quad (9.63)$$

where $\Phi_c \equiv GM_c/(\psi_0 r_c)$ is the gravitational potential at the surface of the body. As long as $m \ll 1$ we have $M_\psi^2 \sim R_0/(3m(R_0))$ and $R \sim 8\pi G\rho$, so that $M_\psi^2 r_c^2 \sim \Phi_c/m(R_0)$. Hence $M_\psi r_c \ll 1$ under the condition (9.63), as expected.

From Eq. (9.61) we find that $\gamma \simeq 1/2$ for the distance r close to r_c (i.e. in the case $M_\psi r \ll 1$). This is not compatible with the experimental bound of γ given in Eq. (8.171). Hence the $f(R)$ gravity models with the light scalaron mass ($M_\psi r_c \ll 1$) do not satisfy local gravity constraints. We caution that in the region of high-density where the condition $\delta R \ll R_0$ is violated the above linear analysis

is no longer valid. In fact this happens for compact objects such as the Earth or the Sun. The mean density of the Earth or the Sun is of the order of $\rho \simeq 1\text{--}10 \text{ g/cm}^3$, which is much larger than the present cosmological density $\rho_c^{(0)} \simeq 10^{-29} \text{ g/cm}^3$. In the environment of such a high density, the field mass M_ψ becomes large such that $M_\psi r_c \gg 1$. This is a non-linear regime in which δR exceeds the background value R_0 . The effect of the chameleon mechanism (see Section 8.4) becomes crucially important in this regime [367, 368, 378, 379, 380, 381]. It is possible to satisfy local gravity constraints for the $f(R)$ models that are designed to have a large scalaron mass in the region of high density.

In order to discuss the chameleon mechanism in $f(R)$ gravity, it is convenient to transform the action (9.1) to the Einstein frame action via the conformal transformation:

$$\tilde{g}_{\mu\nu} = \Omega^2 g_{\mu\nu}, \quad (9.64)$$

where a tilde represents quantities in the Einstein frame. The relation between the Ricci scalars in the two frames is

$$R = \Omega^2 (\tilde{R} + 6\tilde{\square}\omega - 6\tilde{g}^{\mu\nu}\omega_{,\mu}\omega_{,\nu}), \quad (9.65)$$

where

$$\omega_{,\mu} \equiv \frac{\partial_\mu \Omega}{\Omega}, \quad \tilde{\square}\omega \equiv \frac{1}{\sqrt{-\tilde{g}}}\partial_\mu(\sqrt{-\tilde{g}}\tilde{g}^{\mu\nu}\partial_\nu\omega). \quad (9.66)$$

The action (9.1) can be written as

$$S = \int d^4x \sqrt{-g} \left(\frac{1}{2\kappa^2} F R - U \right) + S_m(g_{\mu\nu}, \Psi_m), \quad (9.67)$$

where

$$U = \frac{RF - f}{2\kappa^2}. \quad (9.68)$$

Using Eq. (9.65) and the relation $\sqrt{-g} = \Omega^{-4}\sqrt{-\tilde{g}}$, the action (9.67) is transformed to be

$$S = \int d^4x \sqrt{-\tilde{g}} \left[\frac{1}{2\kappa^2} F \Omega^{-2} (\tilde{R} + 6\tilde{\square}\omega - 6\tilde{g}^{\mu\nu}\omega_{,\mu}\omega_{,\nu}) - \Omega^{-4} U \right] + S_m(g_{\mu\nu}, \Psi_m). \quad (9.69)$$

We obtain a linear action in $\tilde{R}$ for the choice

$$\Omega^2 = F. \quad (9.70)$$

We also introduce a new scalar field ϕ defined by

$$\kappa\phi \equiv \sqrt{\frac{3}{2}} \ln F. \quad (9.71)$$

Since $\Omega = \sqrt{F}$ and $\omega_{,\mu} = \Omega_{,\mu}/\Omega$, it follows that $\omega_{,\mu} = (1/\sqrt{6})\kappa\phi_{,\mu}$. The integral $\int d^4x \sqrt{-\tilde{g}} \tilde{\square}\omega$ vanishes on account of Gauss's theorem by using Eq. (9.66). Then the action in the Einstein frame is

$$S_E = \int d^4x \sqrt{-\tilde{g}} \left[\frac{1}{2\kappa^2} \tilde{R} - \frac{1}{2} \tilde{g}^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - V(\phi) \right] + S_m(g_{\mu\nu}, \Psi_m), \quad (9.72)$$

where

$$V(\phi) = \frac{RF - f}{2\kappa^2 F^2}. \quad (9.73)$$

In the rest of this subsection we shall use the unit $\kappa^2 = 1$. In the Einstein frame the scalar field ϕ has a direct coupling Q with non-relativistic matter. In Section 8.4 we have seen that this coupling has a relation

$$\Omega^2 = F = e^{-2Q\phi}, \quad (9.74)$$

with the conformal factor. From Eq. (9.71) we then find that

$$Q = -1/\sqrt{6}. \quad (9.75)$$

In the absence of the field potential $V(\phi)$ it is not possible to satisfy local gravity constraints because the field propagates freely with a large coupling Q whose strength is of the order of unity [357]. Since a potential $V(\phi)$ with a gravitational origin is present in $f(R)$ gravity, local gravity tests can be escaped through the chameleon mechanism [321, 322], provided that the form of $f(R)$ is appropriately chosen.

The action (9.72) is the same as (8.136) with the identifications $g_{\mu\nu}^{(i)} = g_{\mu\nu}$ and $Q_i = Q$, apart from the difference that a tilde is used in (9.72) for the quantities in the Einstein frame. In $f(R)$ gravity the field ϕ couples to non-relativistic matter universally with the same coupling. One can apply the discussions of Eqs. (8.139)–(8.155) to $f(R)$ gravity just by the replacement $Q \rightarrow -1/\sqrt{6}$ and $r \rightarrow \tilde{r}$. The first and second derivatives of the potential $V(\phi)$ in terms of the field ϕ are given by

$$V_{,\phi} = \sqrt{\frac{2}{3}} \frac{2f - Rf_{,R}}{2f_{,R}^2}, \quad V_{,\phi\phi} = \frac{1}{3f_{,RR}} \left(1 + \frac{Rf_{,RR}}{f_{,R}} - \frac{4ff_{,RR}}{f_{,R}^2} \right). \quad (9.76)$$

Recall that the effective potential is $V_{\text{eff}}(\phi) = V(\phi) + e^{Q\phi}\rho$, where ρ is a conserved density in the Einstein frame. In $f(R)$ gravity the coupling Q is negative so that the effective potential has a local minimum for $V_{,\phi} > 0$, i.e. for $2f > Rf_{,R}$. For

the $f(R)$ models in which the deviation from the Λ CDM model is not significant we have $Rf_{,RR}/f_{,R} \ll 1$ and $4ff_{,RR}/f_{,R}^2 \ll 1$, which gives $V_{,\phi\phi} \simeq 1/(3f_{,RR})$. In fact on a local minimum of $V(\phi)$ we can write

$$V_{,\phi\phi} = \frac{1-m}{3f_{,RR}}. \quad (9.77)$$

As long as $f_{,RR} > 0$ and $m < 1$ the mass squared of the effective potential is positive.

Let us consider local gravity constraints on the $f(R)$ models given in Eqs. (9.9) and (9.10). In the region of high density where local gravity experiments are carried out ($R \gg R_c$), these models behave as

$$f(R) \simeq R - \mu R_c [1 - (R/R_c)^{-2n}], \quad (9.78)$$

which approaches the Λ CDM model in the limit $R/R_c \rightarrow \infty$. Recall that R_c is roughly the same order as the cosmological Ricci scalar R_0 today for μ and n of the order of unity. For the functional form (9.78) we have the following relations

$$F = e^{2\phi/\sqrt{6}} = 1 - 2n\mu(R/R_c)^{-(2n+1)}, \quad (9.79)$$

$$V_{\text{eff}}(\phi) \simeq \frac{1}{2}\mu R_c e^{-4\phi/\sqrt{6}} \left[1 - (2n+1) \left(\frac{-\phi}{\sqrt{6}n\mu} \right)^{2n/(2n+1)} \right] + \rho e^{-\phi/\sqrt{6}}, \quad (9.80)$$

where the effective potential $V_{\text{eff}}(\phi)$ is defined in Eq. (8.143). Inside a spherically symmetric body with a constant energy density ρ_A , the effective potential (9.80) has a minimum at

$$\phi_A \simeq -\sqrt{6}n\mu(R_c/\rho_A)^{2n+1}. \quad (9.81)$$

In the region of high density ($\rho_A \gg R_c \sim H_0^2 \sim \rho_c^{(0)} \sim 10^{-29} \text{ g/cm}^3$), $|\phi_A|$ is very much smaller than 1. Since the quantity $F(R)$ is close to unity from Eq. (9.79), the deviation from the Λ CDM model is small in this regime ($R \gg R_c$). The mass squared about the potential minimum is given by

$$m_A^2 \equiv \frac{d^2 V_{\text{eff}}}{d\phi^2}(\phi_A) \simeq \frac{1}{6n(n+1)\mu} R_c \left(\frac{\rho_A}{R_c} \right)^{2(n+1)}. \quad (9.82)$$

In the region of high density the mass m_A is much larger than the present Hubble parameter H_0 ($\sim \sqrt{R_c}$) for n and μ of the order of unity. The density ρ_B outside the body is generally much smaller than the density ρ_A inside the body. In this case the effective potential (9.80) has a minimum at $\phi_B \simeq -\sqrt{6}n\mu(R_c/\rho_B)^{2n+1}$ with a mass squared $m_B^2 \equiv \frac{d^2 V_{\text{eff}}}{d\phi^2}(\phi_B) \ll m_A^2$.

In the region $R \gg R_c$, the quantity m behaves as

$$m \simeq 2n(2n+1)\mu(R_c/R)^{2n+1}. \quad (9.83)$$

Since the models are close to the Λ CDM model for $R \gg R_c$ one can approximate $R \simeq \rho$ in this region. Then the field value ϕ_B is estimated as

$$\phi_B \simeq -\frac{\sqrt{6}}{2(2n+1)}m(R_B), \quad (9.84)$$

where R_B is the Ricci scalar outside the body. Taking note of the relation $|\phi_A| \ll |\phi_B|$ ($\ll 1$), we find that the thin-shell parameter defined in Eq. (8.158) can be estimated as

$$\epsilon_{\text{th}} \simeq \frac{1}{2(2n+1)} \frac{m(R_B)}{\Phi_c}. \quad (9.85)$$

The models can be consistent with local gravity constraints if ϵ_{th} is much smaller than unity. For n of the order of unity this translates into the condition

$$m(R_B) \ll \Phi_c. \quad (9.86)$$

Since $\Phi_\odot \simeq 2.1 \times 10^{-6}$ and $\Phi_\oplus \simeq 7.0 \times 10^{-10}$ for Sun and Earth, respectively, the deviation parameter m is required to be very much smaller than unity in the region outside the body. Note that the condition (9.86) is opposite to the condition (9.63) required for the validity of the linear expansion of R about the background value R_0 . The region in which the chameleon mechanism works corresponds to the non-linear regime in which such a linear expansion of R is no longer valid.

Let us consider constraints on model parameters by using the bound coming from the violation of equivalence principle. Using the relation $|\phi_{B,\oplus}| = \sqrt{6}n\mu(R_c/\rho_B)^{2n+1}$, the bound (8.173) translates into

$$\frac{n\mu}{x_1^{2n+1}} \left(\frac{R_1}{\rho_B} \right)^{2n+1} < 1.5 \times 10^{-15}, \quad (9.87)$$

where x_1 is defined by $x_1 \equiv R_1/R_c$. Let us consider the case in which the Lagrangian density is given by (9.78) for $R \geq R_1$. If we use the original models of Hu and Sawicki [368] and Starobinsky [369], then there are some modifications for the estimation of R_1 , but this change is not significant when we place constraints on model parameters.

The de Sitter point for the model (9.78) corresponds to $\mu = x_1^{2n+1}/[2(x_1^{2n} - n - 1)]$. Substituting this relation into Eq. (9.87), we find

$$\frac{n}{2(x_1^{2n} - n - 1)} \left(\frac{R_1}{\rho_B} \right)^{2n+1} < 1.5 \times 10^{-15}. \quad (9.88)$$

For the stability of the de Sitter point we require that $m(R_1) < 1$, which translates into the condition $x_1^{2n} > 2n^2 + 3n + 1$. Hence the term $n/[2(x_1^{2n} - n - 1)]$ in Eq. (9.88) is smaller than 0.25 for $n > 0$.

We now assume that R_1 and ρ_B are of the orders of the present cosmological density 10^{-29} g/cm^3 and the baryonic/dark matter density 10^{-24} g/cm^3 in our galaxy, respectively. From Eq. (9.88) we obtain the constraint

$$n > 0.9. \quad (9.89)$$

Thus n does not need to be much larger than unity. Under the condition (9.89) one can see an appreciable deviation from the Λ CDM model cosmologically ($m \lesssim \mathcal{O}(0.1)$) as R decreases to the order of R_c , while satisfying the condition (9.86) in the region of high density.

Thus we have seen that the models (9.9) and (9.10) are carefully constructed to satisfy local gravity constraints even when n is of the order of 1. Meanwhile, other $f(R)$ models are likely to be plagued by the problem of tuning model parameters. For example, let us also study the following model [363, 376]

$$f(R) = R - \mu R_c (R/R_c)^p, \quad (0 < p < 1, \mu, R_c > 0). \quad (9.90)$$

The model with negative p is excluded because it corresponds to $f_{,RR} < 0$. Since the field ϕ_B is given by $\phi_B = -(\sqrt{6}/2)\mu p (R_c/\rho_B)^{1-p}$, the experimental bound (8.173) translates into

$$\mu p \left(\frac{R_c}{\rho_B} \right)^{1-p} < 1.5 \times 10^{-15}. \quad (9.91)$$

The de Sitter point, $x_1 = R_1/R_c$, satisfies the relation $\mu = x_1^{1-p}/(2-p)$. Hence the bound (9.91) yields

$$\frac{p}{2-p} \left(\frac{R_1}{\rho_B} \right)^{1-p} < 1.5 \times 10^{-15}. \quad (9.92)$$

Taking $R_1 = 10^{-29} \text{ g/cm}^3$ and $\rho_B = 10^{-24} \text{ g/cm}^3$, we obtain the constraint

$$p < 3 \times 10^{-10}. \quad (9.93)$$

This means that the deviation from the Λ CDM model is very small. Since the power $(1-p)$ does not exceed unity in Eq. (9.92), we need to choose values of p much smaller than unity to satisfy the condition (9.92). On the other hand, the model (9.78) can easily make $|\phi_B|$ very small due to the presence of a larger power $2n+1$.

In summary we have shown that the models (9.9) and (9.10) can be consistent with local gravity constraints for $n \gtrsim 1$, whereas in the model (9.8) we require that $p < 10^{-10}$. The model (9.11) can be regarded as the special case of the models

(9.9) and (9.10) with $n \gg 1$. The models (9.9), (9.10), and (9.11) are very close to the Λ CDM model for $R \gg R_c$, but it is possible to see the deviation from it as R approaches R_c . As we see from Eq. (9.79), the field value $\phi = 0$ corresponds to the limit $R \rightarrow \infty$ with a finite effective potential for the models (9.9) and (9.10) [382]. As we go back to the past, the field value ϕ_A at the potential minimum approaches $\phi = 0$. The perturbations in ϕ need to be suppressed relative to its background value in order to avoid reaching the curvature singularity at $\phi = 0$. This is related to the suppression of the oscillating mode in ϕ with the large mass $M_\phi^2 \simeq f_{,R}/(3f_{,RR})$ [369, 371].

In the strong gravitational background (such as neutron stars), Ref. [383] showed that for the model (9.10) it is difficult to obtain thin-shell solutions inside a spherically symmetric body with constant density. For chameleon models with general couplings Q , a thin-shell field profile has been analytically derived in Ref. [324] by employing a linear expansion in terms of the gravitational potential Φ_c at the surface of a compact object with constant density. Using the boundary condition set by analytic solutions, Ref. [324] also numerically confirmed the existence of thin-shell solutions for $\Phi_c \lesssim 0.3$ in the case of inverse power-law potentials $V(\phi) = M^{4+n}\phi^{-n}$. The effect of the relativistic pressure is important around the center of the body so that the field tends to roll down the potential quickly unless the boundary condition is carefully chosen. Note that realistic stars have densities $\rho_A(r)$ that globally decrease as a function of r . The numerical simulation of Refs. [384, 385] showed that thin-shell solutions are present for the $f(R)$ model (9.10) by considering a polytropic equation of state even in the strong gravitational background.

9.1.4 The viability of $f(R)$ gravity models in the Palatini formalism

In $f(R)$ theory of the Palatini formalism the field equation (9.5) is of the second order, unlike the fourth-order equation (9.3) in the metric case. Hence the scalar-field degree of freedom does not have a dynamical evolution as in the case of General Relativity. Recall that in the metric formalism the condition $f_{,RR} > 0$ is required in order to avoid the instability problem of perturbations associated with the field mass squared M_ψ^2 . In the Palatini formalism there are no such restrictions because of the absence of the dynamical degree of freedom for the field. This is one of the reasons why the viable background cosmological evolution can be realized fairly easily in the Palatini case, as we will see below.

Let us discuss the viability of $f(R)$ gravity models in the Palatini formalism [386, 387, 388, 389, 390, 391]. We first study the cosmological dynamics in this formalism. Taking into account non-relativistic matter and radiation, Eqs. (9.4) and

(9.5) give the following equations of motion in the flat FLRW background:

$$6F \left(H + \dot{F}/2F \right)^2 - f = \kappa^2(\rho_m + 2\rho_r), \quad (9.94)$$

$$FR - 2f = -\kappa^2\rho_m. \quad (9.95)$$

Taking the derivative of Eq. (9.95) with respect to t and using the conservation equation for ρ_m , we find [399]

$$\dot{R} = \frac{3\kappa^2 H \rho_m}{R F_{,R} - F} = -3H \frac{FR - 2f}{F_{,R}R - F}. \quad (9.96)$$

Combining Eqs. (9.94) and (9.96) gives

$$H^2 = \frac{2\kappa^2(\rho_m + \rho_r) + FR - f}{6F\xi}, \quad (9.97)$$

where

$$\xi \equiv \left[1 - \frac{3}{2} \frac{F_{,R}(FR - 2f)}{F(F_{,R}R - F)} \right]^2. \quad (9.98)$$

We introduce the following dimensionless variables

$$y_1 \equiv \frac{FR - f}{6F\xi H^2}, \quad y_2 \equiv \frac{\kappa^2 \rho_r}{3F\xi H^2}. \quad (9.99)$$

In terms of these variables the constraint equation (9.97) becomes

$$\frac{\kappa^2 \rho_m}{3F\xi H^2} = 1 - y_1 - y_2. \quad (9.100)$$

Differentiating Eq. (9.97) we obtain

$$2 \frac{\dot{H}}{H^2} = -3 + 3y_1 - y_2 - \frac{\dot{F}}{HF} - \frac{\dot{\xi}}{H\xi} + \frac{\dot{F}R}{6F\xi H^3}. \quad (9.101)$$

The variables y_1 and y_2 obey the following equations of motion [392]:

$$\frac{dy_1}{dN} = y_1 [3 - 3y_1 + y_2 + C(R)(1 - y_1)], \quad (9.102)$$

$$\frac{dy_2}{dN} = y_2 [-1 - 3y_1 + y_2 - C(R)y_1], \quad (9.103)$$

where

$$C(R) \equiv \frac{R\dot{F}}{H(FR - f)} = -3 \frac{(FR - 2f)F_{,R}R}{(FR - f)(F_{,R}R - F)}. \quad (9.104)$$

The following constraint equation also holds

$$\frac{FR - 2f}{FR - f} = -\frac{1 - y_1 - y_2}{2y_1}, \quad (9.105)$$

which shows that R and thus $C(R)$ can in principle be expressed in terms of the variables y_1 and y_2 .

The behavior of the variables y_1 and y_2 depends on the function $C(R)$. For the well-behaved function $C(R)$, but excluding the cases $C(R) = -3, -4$, we obtain the following fixed points for the system (9.102)–(9.103):

$$(i) P_r : (y_1, y_2) = (0, 1), \quad (ii) P_m : (y_1, y_2) = (0, 0), \quad (iii) P_d : (y_1, y_2) = (1, 0). \quad (9.106)$$

The eigenvalues of the Jacobian matrices for the linearized perturbation equations about the above fixed points are given by

$$(i) P_r : 4 + C(R), 1, \quad (ii) P_m : 3 + C(R), -1, \\ (iii) P_d : -3 - C(R), -4 - C(R). \quad (9.107)$$

The effective equation of state of the system is

$$w_{\text{eff}} = -y_1 + \frac{1}{3}y_2 + \frac{\dot{F}}{3HF} + \frac{\dot{\xi}}{3H\xi} - \frac{\dot{F}R}{18F\xi H^3}. \quad (9.108)$$

For the Λ CDM model, $f(R) = R - 2\Lambda$, we have $C(R) = 0$ and hence the points P_r , P_m , and P_d correspond to radiation ($w_{\text{eff}} = 1/3$, unstable), matter ($w_{\text{eff}} = 0$, saddle), and de Sitter ($w_{\text{eff}} = -1$, stable) points, respectively. Thus it is possible to realize the sequence of radiation, matter, and de Sitter epochs.

Let us next consider the model $f(R) = R - \beta/R^n$ with $\beta > 0$ and $n > -1$. From Eqs. (9.104) and (9.105) this model gives the following relations

$$C(R) = 3n \frac{R^{1+n} - (2+n)\beta}{R^{1+n} + n(2+n)\beta} \quad \text{and} \quad R^{1+n} = \frac{\beta[3y_1 + n(y_1 - y_2 + 1) - y_2 + 1]}{2y_1}. \quad (9.109)$$

We focus on the case in which the condition $\beta/R^{1+n} \ll 1$ is satisfied during the radiation era (i.e. $f(R) \simeq R$). We then find that the points P_r , P_m , P_d behave as radiation, matter, and de Sitter points, respectively, with the eigenvalues

$$(i) P_r : (3n + 4, 1), \quad (ii) P_m : (3n + 3, -1), \quad (iii) P_d : (-3, -4). \quad (9.110)$$

Note that the de Sitter point P_d satisfies the relation, $R^{1+n} = (2+n)\beta$, which exists for $n > -2$ provided that $R > 0$ and $\beta > 0$. If $n > -1$ and $\beta > 0$, the points P_r , P_m , P_d correspond to an unstable node, a saddle, and a stable node, respectively. Numerical simulations in Ref. [392] indeed confirmed that such theories admit the sequence of radiation, matter, and de Sitter phases. This property is different from the $f(R)$ gravity in the metric formalism – the model $f(R) = R - \beta/R^n$ ($\beta > 0, n > 0$) does not give rise to viable cosmological evolution (because the condition $f_{,RR} > 0$ is not satisfied).

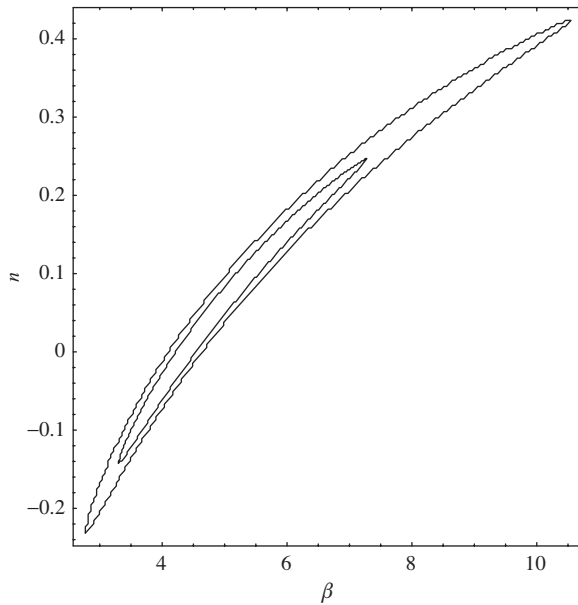


Figure 9.2 The observational contours at the 68% and 95% confidence levels for the model based on the theory $f(R) = R - \beta/R^n$ in the Palatini formalism, constrained by SNLS, BAO, and CMB data. From Ref. [392].

In Refs. [393, 392] observational constraints on the model $f(R) = R - \beta/R^n$ have been derived by using the SN Ia, BAO, and CMB data. As we show in Fig. 9.2 the two parameters n and β are constrained to be $n \in [-0.23, 0.42]$ and $\beta \in [2.73, 10.6]$ at the 95% confidence level (the unit $H_0 = 1$ is chosen for the normalization of β). Since the allowed values of n are close to 0, the above model is not particularly favored over the Λ CDM model.

In Section 9.2 we will show that $f(R)$ theory in the Palatini formalism is equivalent to Brans–Dicke theory [63] with a parameter $\omega_{\text{BD}} = -3/2$ (and with a field potential) [360, 387, 391]. If we transform the action (9.121) of the Brans–Dicke theory with a field potential by a conformal transformation, the resulting Einstein frame action corresponds to a coupled quintessence scenario with a constant coupling Q between the field ϕ and non-relativistic matter [273]. From the relation (9.122) between ω_{BD} and Q we see that $f(R)$ theory in the Palatini formalism ($\omega_{\text{BD}} = -3/2$) corresponds to the infinite coupling, $Q^2 \rightarrow \infty$. It is expected that this large coupling should lead to some observable effects.

In Ref. [394] the equations for matter density perturbations have been derived in the Jordan frame of $f(R)$ theory in the Palatini formalism. Under the sub-horizon approximation in which the comoving wavenumber k is much larger than aH , the

equation for matter perturbations δ_m is approximately given by [395]

$$\ddot{\delta}_m + 2H\dot{\delta}_m - \frac{\rho_m}{2F} \left(1 + \frac{m}{1-m} \frac{k^2}{a^2 R} \right) \delta_m \simeq 0, \quad (9.111)$$

where $m = Rf_{,RR}/f_{,R}$ is the deviation parameter from the Λ CDM model. The second term in the bracket of Eq. (9.111) plays a crucial role when we discuss the evolution of perturbations. Since R is of the order of H^2 we have that $k^2/(a^2 R) \sim (k/aH)^2 \gg 1$ for sub-horizon modes. For smaller scale modes (i.e. larger k) the evolution of δ_m is quite different from that in the Λ CDM model. If $m > 0$ there is violent growth of matter perturbations, whereas if $m < 0$ there is a damped oscillation of δ_m . The constraint from the large-scale structure gives the bound $|m| \lesssim 10^{-5} - 10^{-4}$ even at the present epoch irrespective of the form of $f(R)$ [395]. Hence the $f(R)$ models in the Palatini formalism are constrained to be very close to the Λ CDM model from the evolution of matter perturbations. In $f(R)$ theory in the metric formalism the deviation parameter m can grow as large as the order of 0.1 around the present epoch, e.g., for the models (9.9)–(9.11), which means that the models can show appreciable deviation from the Λ CDM model.

The $f(R)$ theory in the Palatini formalism also gives rise to non-perturbative corrections to the matter action as well as strong couplings between gravity and matter. This problem was first raised by Flanagan [396] using Dirac particles for the matter action and was studied again for the matter action of a Higgs field [397]. Starting from the Brans–Dicke action (9.121) with $\omega_{\text{BD}} = -3/2$ along the line of Ref. [398] by taking into account a Higgs boson h with mass m_h , one can show that, for the model $f(R) = R - \beta/R$, a perturbative expansion for the Brans–Dicke field ϕ leads to large non-perturbative corrections to a local matter action [399]. As long as the Palatini $f(R)$ models are designed to explain the late-time cosmic acceleration, they are in conflict with the standard model of particle physics unless the deviation from the Λ CDM model is very small. This is associated with the infinitely large coupling problem mentioned above.

9.2 Scalar-tensor theories

As we have seen in this section and in the previous chapters, most models of dark energy rely on scalar fields. Scalar fields have a long history in cosmology, starting from Brans–Dicke theory [63] in which gravity is mediated by a scalar field in addition to the metric tensor field (see the book [400]). Brans–Dicke theory was an attempt to revive Mach’s principle (according to which inertia arises when a body is accelerated with respect to the global mass distribution in the Universe) by linking the gravitational constant to a cosmic field. At the same time, Brans–Dicke theory incorporated Dirac’s suggestion that G varies in time in order to

explain the coincidence that combinations like $(\hbar^2 H_0 c^5 / G)^{1/3} \simeq 68 h^{1/3} \text{ MeV}$ or $(e^4 H_0 c^3 / G)^{1/3} \simeq 2.5 h^{1/3} \text{ MeV}$ are of the order of typical particle masses.

Brans–Dicke theory is just a particular example of scalar-tensor theories. These are probably the simplest example of modified gravity models and as such one of the most intensely studied alternatives to General Relativity. After the discovery of cosmic acceleration, they have been invoked by several authors [56, 57, 58, 59, 60, 374, 401, 402] to generalize the cosmological constant and to explain the fine-tuning and the coincidence problem. In this section we discuss their properties as dark energy candidates.

The action for scalar-tensor theories is given by

$$S = \int d^4x \sqrt{-g} \left[\frac{1}{2} f(\varphi, R) - \frac{1}{2} \zeta(\varphi) (\nabla\varphi)^2 \right] + S_m(g_{\mu\nu}, \Psi_m), \quad (9.112)$$

where f is a general function of the scalar field φ and the Ricci scalar R , ζ is a function of φ , and S_m is a matter Lagrangian that depends on the metric $g_{\mu\nu}$ and matter fields Ψ_m . We choose units such that $\kappa^2 = 1$.

The action (9.112) includes a wide variety of theories such as $f(R)$ gravity, Brans–Dicke theory, and dilaton gravity. The $f(R)$ gravity corresponds to the choice $f(\varphi, R) = f(R)$ and $\zeta = 0$. The action in Brans–Dicke theory is $f = \varphi R$ and $\zeta = \omega_{\text{BD}}/\varphi$, where ω_{BD} is called the Brans–Dicke parameter [63]. One can generalize Brans–Dicke theory by adding the field potential $U(\varphi)$ to the original action, i.e. $f = \varphi R - 2U(\varphi)$ and $\zeta = \omega_{\text{BD}}/\varphi$. The dilaton gravity arising from low-energy effective string theory [64] corresponds to $f = 2e^{-\varphi} R - 2U(\varphi)$ and $\zeta(\varphi) = -2e^{-\varphi}$, where we have introduced the dilaton potential $U(\varphi)$. The action (9.112) can be transformed to the Einstein frame under the conformal transformation (9.64) with the choice

$$\Omega^2 = F \equiv \frac{\partial f}{\partial R}, \quad (9.113)$$

where F is positive in order to ensure that gravity is attractive.

Let us consider theories of the type

$$f(\varphi, R) = F(\varphi)R - 2U(\varphi), \quad (9.114)$$

for which the conformal factor Ω depends on φ only. Following a procedure similar to the one we employed for $f(R)$ gravity, we obtain the action in the Einstein frame:

$$S_E = \int d^4x \sqrt{-\tilde{g}} \left[\frac{1}{2} \tilde{R} - \frac{1}{2} (\tilde{\nabla}\phi)^2 - V(\phi) \right] + S_m(\tilde{g}_{\mu\nu} F^{-1}, \Psi_m), \quad (9.115)$$

where we have introduced a new scalar field ϕ in order to make the kinetic term canonical:

$$\phi = \int d\varphi \sqrt{\frac{3}{2} \left(\frac{F_{,\varphi}}{F} \right)^2 + \frac{\zeta}{F}}. \quad (9.116)$$

The potential $V(\phi)$ is given by

$$V = U/F^2. \quad (9.117)$$

Recall that the Einstein frame action (9.72) in $f(R)$ gravity in the metric formalism is the same as Eq. (9.115) with the correspondence $\phi = (\sqrt{6}/2) \ln F$ and $V = (RF - f)/2F^2$. From Eq. (9.117) the potential U in the Jordan frame is given by $U = (RF - f)/2$.

In order to describe the strength of the coupling between dark energy and non-relativistic matter, we introduce the following quantity

$$Q \equiv -\frac{F_{,\phi}}{2F} = -\frac{F_{,\phi}}{F} \left[\frac{3}{2} \left(\frac{F_{,\phi}}{F} \right)^2 + \frac{\zeta}{F} \right]^{-1/2}. \quad (9.118)$$

Since $F = e^{2\phi/\sqrt{6}}$ in $f(R)$ gravity in the metric formalism we have that $Q = -1/\sqrt{6}$. If Q is a constant, the following relations hold from Eqs. (9.116) and (9.118):

$$F = e^{-2Q\phi}, \quad \zeta = (1 - 6Q^2)F \left(\frac{d\phi}{d\varphi} \right)^2. \quad (9.119)$$

Then the action (9.112) in the Jordan frame yields [273]

$$S = \int d^4x \sqrt{-g} \left[\frac{1}{2} F(\phi) R - \frac{1}{2} (1 - 6Q^2) F(\phi) (\nabla\phi)^2 - U(\phi) \right] + S_m(g_{\mu\nu}, \Psi_m). \quad (9.120)$$

In the limit $Q \rightarrow 0$ the action (9.120) reduces to the one for a minimally coupled scalar field ϕ with the potential $U(\phi)$. The transformation of the Jordan frame action (9.120) via a conformal transformation $\tilde{g}_{\mu\nu} = F(\phi)g_{\mu\nu}$ gives rise to the Einstein frame action (9.115) with a constant coupling Q , which is equivalent to the action (8.136) for the chameleon scalar field.

It is instructive to compare (9.120) with the action of Brans–Dicke theory with a potential U :

$$S = \int d^4x \sqrt{-g} \left[\frac{1}{2} \psi R - \frac{\omega_{\text{BD}}}{2\psi} (\nabla\psi)^2 - U(\psi) \right] + S_m(g_{\mu\nu}, \Psi_m). \quad (9.121)$$

Setting $\psi = F = e^{-2Q\phi}$, one easily finds that the two actions are equivalent if the parameter ω_{BD} is related to Q via the relation [273]

$$3 + 2\omega_{\text{BD}} = \frac{1}{2Q^2}. \quad (9.122)$$

Under this condition, the theories given by (9.120) are equivalent to Brans–Dicke theory with a potential U . In the General Relativistic limit, $Q \rightarrow 0$, we have $\omega_{\text{BD}} \rightarrow \infty$ as expected.

Taking the variation of Eq. (9.121) with respect to $g_{\mu\nu}$ and ψ leads to the following equations

$$R_{\mu\nu}(g) - \frac{1}{2}g_{\mu\nu}R(g) = \frac{1}{\psi}T_{\mu\nu} - \frac{1}{\psi}g_{\mu\nu}U(\psi) + \frac{1}{\psi}(\nabla_\mu \nabla_\nu \psi - g_{\mu\nu}\square\psi) + \frac{\omega_{\text{BD}}}{\psi^2} \left[\partial_\mu \psi \partial_\nu \psi - \frac{1}{2}g_{\mu\nu}(\nabla\psi)^2 \right], \quad (9.123)$$

$$(3 + 2\omega_{\text{BD}})\square\psi + 4U(\psi) - 2\psi U_{,\psi} = T. \quad (9.124)$$

In order to find the relation of Brans–Dicke theory with the $f(R)$ theories, let us consider the following correspondence

$$\psi = F(R), \quad U(\psi) = \frac{1}{2} [R(\psi)F - f(R(\psi))], \quad (9.125)$$

where $R = R(g)$ in the metric case and $R = R(T)$ in the Palatini case. Comparing Eqs. (9.123) and (9.124) with Eqs. (9.2) and (9.3), it can readily be seen that $f(R)$ theory in the metric formalism corresponds to the above generalized Brans–Dicke theory with $\omega_{\text{BD}} = 0$. Similarly the comparison of Eqs. (9.123) and (9.124) with Eqs. (9.6) and (9.5) shows that $f(R)$ theory in the Palatini formalism corresponds to the generalized Brans–Dicke theory with $\omega_{\text{BD}} = -3/2$. Since the coupling Q for the $f(R)$ theory in the metric formalism is given by $Q = -1/\sqrt{6}$, it also follows from Eq. (9.122) that the Brans–Dicke parameter ω_{BD} is equivalent to 0. From Eq. (9.122) $f(R)$ theory in the Palatini formalism ($\omega_{\text{BD}} = -3/2$) corresponds to the infinite matter coupling, $Q^2 \rightarrow \infty$.

In dilaton gravity [$f(\varphi, R) = 2e^{-\varphi}R - 2U$ and $\zeta = -2e^{-\varphi}$] one has $F = 2e^{-\varphi}$ and $\phi = \varphi/\sqrt{2}$. In this case we find from Eqs. (9.118) and (9.122) that the coupling is also constant ($Q = 1/\sqrt{2}$) with the Brans–Dicke parameter $\omega_{\text{BD}} = -1$. The above discussion shows that the action (9.120) with $F(\phi) = e^{-2Q\phi}$ corresponds to Brans–Dicke theory with the potential $U(\phi)$, which includes a wide variety of theories such as $f(R)$ theories in the metric and Palatini formalisms, and dilaton gravity.

There are theories that give rise to varying couplings Q . For example a non-minimally coupled scalar field with a coupling ξ corresponds to the choice $F(\varphi) = 1 - \xi\varphi^2$ and $\zeta(\varphi) = 1$. In this case the coupling $Q(\varphi)$ is field-dependent:

$$Q(\varphi) = \frac{\xi\varphi}{[1 - \xi\varphi^2(1 - 6\xi)]^{1/2}}. \quad (9.126)$$

Note that $Q \simeq \xi\varphi$ for $|\xi| \ll 1$ and $Q \simeq \pm 1/\sqrt{6}$ in the limit $|\xi| \gg 1$.

In the following we shall focus on the constant coupling models based on the action (9.120) with $F(\phi) = e^{-2Q\phi}$.

9.2.1 Cosmological dynamics

We shall study the cosmological dynamics for the Jordan frame action (9.120) with $F(\phi) = e^{-2Q\phi}$ in the presence of a non-relativistic fluid with energy density ρ_m and a radiation fluid with energy density ρ_r . We regard here the Jordan frame as a “physical” one due to the usual conservation of non-relativistic matter ($\rho_m \propto a^{-3}$) in this frame. In the Einstein frame the system is described by a coupled quintessence scenario with the potential $V = U/F^2$. One can study the cosmological dynamics in the Einstein frame and transform back to the Jordan frame, but we shall carry out the analysis directly in the Jordan frame.

In the flat FLRW background the variation of the action (9.120) with respect to $g_{\mu\nu}$ and ϕ leads to the following equations of motion

$$3FH^2 = \frac{1}{2}(1 - 6Q^2)F\dot{\phi}^2 + U - 3H\dot{F} + \rho_m + \rho_r, \quad (9.127)$$

$$2F\dot{H} = -(1 - 6Q^2)F\dot{\phi}^2 - \ddot{F} + H\dot{F} - \rho_m - (4/3)\rho_r, \quad (9.128)$$

$$(1 - 6Q^2)F[\ddot{\phi} + 3H\dot{\phi} + (\dot{F}/2F)\dot{\phi}] + U_{,\phi} + QFR = 0. \quad (9.129)$$

We introduce the following variables

$$x_1 \equiv \frac{\dot{\phi}}{\sqrt{6}H}, \quad x_2 \equiv \frac{1}{H}\sqrt{\frac{U}{3F}}, \quad x_3 \equiv \frac{1}{H}\sqrt{\frac{\rho_r}{3F}}, \quad (9.130)$$

and

$$\Omega_m \equiv \frac{\rho_m}{3FH^2}, \quad \Omega_r \equiv x_3^2, \quad \Omega_{DE} \equiv (1 - 6Q^2)x_1^2 + x_2^2 + 2\sqrt{6}Qx_1. \quad (9.131)$$

These satisfy the relation $\Omega_m + \Omega_r + \Omega_{DE} = 1$ from Eq. (9.127).

Using Eqs. (9.127)–(9.129), we obtain the differential equations for x_1 , x_2 , and x_3 :

$$\begin{aligned} \frac{dx_1}{dN} &= \frac{\sqrt{6}}{2}(\lambda x_2^2 - \sqrt{6}x_1) + \frac{\sqrt{6}Q}{2} \left[(5 - 6Q^2)x_1^2 + 2\sqrt{6}Qx_1 - 3x_2^2 + x_3^2 - 1 \right] \\ &\quad - x_1 \frac{\dot{H}}{H^2}, \end{aligned} \quad (9.132)$$

$$\frac{dx_2}{dN} = \frac{\sqrt{6}}{2}(2Q - \lambda)x_1x_2 - x_2 \frac{\dot{H}}{H^2}, \quad (9.133)$$

$$\frac{dx_3}{dN} = \sqrt{6}Qx_1x_3 - 2x_3 - x_3 \frac{\dot{H}}{H^2}, \quad (9.134)$$

where

$$\lambda \equiv -U_{,\phi}/U, \quad (9.135)$$

and $\dot{H}/H^2$ is given by

$$\begin{aligned} \frac{\dot{H}}{H^2} = & -\frac{1-6Q^2}{2} \left(3 + 3x_1^2 - 3x_2^2 + x_3^2 - 6Q^2x_1^2 + 2\sqrt{6}Qx_1 \right) \\ & + 3Q(\lambda x_2^2 - 4Q). \end{aligned} \quad (9.136)$$

The effective equation of state of the system is

$$\begin{aligned} w_{\text{eff}} = & -1 + \frac{1-6Q^2}{3} (3 + 3x_1^2 - 3x_2^2 + x_3^2 - 6Q^2x_1^2 + 2\sqrt{6}Qx_1) \\ & - 2Q(\lambda x_2^2 - 4Q). \end{aligned} \quad (9.137)$$

In the absence of radiation ($x_3 = 0$), the fixed points of the system (9.132)–(9.134) for constant λ are given by [273]

- (a) ϕ MDE

$$(x_1, x_2) = \left(\frac{\sqrt{6}Q}{3(2Q^2 - 1)}, 0 \right), \quad \Omega_m = \frac{3 - 2Q^2}{3(1 - 2Q^2)^2}, \quad w_{\text{eff}} = \frac{4Q^2}{3(1 - 2Q^2)}. \quad (9.138)$$

- (b) Kinetic points

$$(x_1, x_2) = \left(\frac{1}{\sqrt{6}Q \pm 1}, 0 \right), \quad \Omega_m = 0, \quad w_{\text{eff}} = \frac{3 \mp \sqrt{6}Q}{3(1 \pm \sqrt{6}Q)}. \quad (9.139)$$

- (c) Scalar-field dominated point

$$\begin{aligned} (x_1, x_2) = & \left(\frac{\sqrt{6}(4Q - \lambda)}{6(4Q^2 - Q\lambda - 1)}, \left[\frac{6 - \lambda^2 + 8Q\lambda - 16Q^2}{6(4Q^2 - Q\lambda - 1)^2} \right]^{1/2} \right), \quad \Omega_m = 0, \\ w_{\text{eff}} = & -\frac{20Q^2 - 9Q\lambda - 3 + \lambda^2}{3(4Q^2 - Q\lambda - 1)}. \end{aligned} \quad (9.140)$$

- (d) Scaling solution

$$\begin{aligned} (x_1, x_2) = & \left(\frac{\sqrt{6}}{2\lambda}, \sqrt{\frac{3 + 2Q\lambda - 6Q^2}{2\lambda^2}} \right), \quad \Omega_m = 1 - \frac{3 - 12Q^2 + 7Q\lambda}{\lambda^2}, \\ w_{\text{eff}} = & -\frac{2Q}{\lambda}. \end{aligned} \quad (9.141)$$

- (e) de Sitter point (present for $\lambda = 4Q$)

$$(x_1, x_2) = (0, 1), \quad \Omega_m = 0, \quad w_{\text{eff}} = -1. \quad (9.142)$$

One can confirm that the de Sitter point (e) exists for $\lambda = 4Q$, by setting $\dot{\phi} = 0$ in Eqs. (9.127)–(9.129). This is the special case of the scalar-field dominated point (c).

(I) Constant λ

Let us study the case of non-zero values of Q with constant λ , i.e. for the exponential potential $U(\phi) = U_0 e^{-\lambda\phi}$. We do not consider the special case of $\lambda = 4Q$. The stability of the fixed points (a)–(e) can be found as usual by considering the eigenvalues of the 2×2 Jacobian matrix of perturbations, see the problem 9.3. The matter-dominated epoch can be realized either by the point (a) or by the point (d). If the point (a) is responsible for the matter era, we require the condition $Q^2 \ll 1$. We then have $\Omega_m \simeq 1 + 10Q^2/3 > 1$ and $w_{\text{eff}} \simeq 4Q^2/3$. When $Q^2 \ll 1$ the scalar-field dominated point (c) yields an accelerated expansion of the Universe provided that $-\sqrt{2} + 4Q < \lambda < \sqrt{2} + 4Q$. Under these conditions the point (a) is followed by the late-time cosmic acceleration. The scaling solution (d) can give rise to the equation of state, $w_{\text{eff}} \simeq 0$ for $|Q| \ll |\lambda|$. In this case, however, the condition $w_{\text{eff}} < -1/3$ for the point (c) gives $\lambda^2 \lesssim 2$. Then the energy fraction of the pressureless matter for the point (d) does not satisfy the condition $\Omega_m \simeq 1$. From the above discussion the viable cosmological trajectory for constant λ corresponds to the sequence from the point (a) to the scalar-field dominated point (c) under the conditions $Q^2 \ll 1$ and $-\sqrt{2} + 4Q < \lambda < \sqrt{2} + 4Q$. In the Einstein frame this corresponds to the coupled quintessence scenario with the exponential potential $V = U/F^2 = U_0 e^{-(\lambda-4Q)\phi}$ discussed in Section 8.3.

(II) Varying λ

Let us next proceed to the case where λ varies with time. When the time scale of the variation of λ is smaller than that of the cosmic expansion, the fixed points derived above in the case of constant λ can be regarded as the “instantaneous” fixed points. This allows the possibility that the matter era is realized by the point (d) with $|Q| \ll |\lambda|$ and that the solutions finally approach either the de Sitter point (e) with $\lambda = 4Q$ or the accelerated point (c).

In the following let us focus on the case in which the matter solution (d) is followed by the de Sitter solution (e). In order to study the stability of the point (e) we define a variable $x_4 \equiv F$. This satisfies the following equation

$$\frac{dx_4}{dN} = -2\sqrt{6}Qx_1x_4. \quad (9.143)$$

Considering the 3×3 matrix for perturbations δx_1 , δx_2 , and δx_4 around the point (e), we obtain the eigenvalues

$$-3, \quad -\frac{3}{2} \left[1 \pm \sqrt{1 - \frac{8}{3}F_1 Q \frac{d\lambda}{dF}(F_1)} \right], \quad (9.144)$$

where $F_1 \equiv F(\phi_1)$ is the value of F at the de Sitter point with the field value ϕ_1 . Since $F_1 > 0$, we find that the de Sitter point is stable for

$$Q \frac{d\lambda}{dF}(F_1) > 0, \quad \text{i.e.} \quad \frac{d\lambda}{d\phi}(\phi_1) < 0. \quad (9.145)$$

For $f(R)$ theory in the metric formalism ($Q = -1/\sqrt{6}$), this condition translates into $d\lambda/dF < 0$. Since $F = e^{2\phi/\sqrt{6}} = df/dR$ and $U = (RF - f)/2$ in this case, we have $\lambda = -Rf_{,R}/(\sqrt{6}V)$. Then, together with the fact that $Rf_{,R} = 2f$ holds for the de Sitter point, the condition $d\lambda/dF < 0$ is equivalent to $R < f_{,R}/f_{,RR}$. For positive R this gives $0 < Rf_{,RR}/f_{,R} < 1$, which agrees with the condition (9.35).

Let us consider the $f(R)$ model (9.78) in the metric formalism. This recovers the models (9.9) and (9.10) in the regime $R \gg R_c$. Since $e^{2\phi/\sqrt{6}} = 1 - 2n\mu(R/R_c)^{-(2n+1)}$ in this case, the potential $U = (FR - f)/2$ in the Jordan frame yields

$$U(\phi) = \frac{\mu R_c}{2} \left[1 - \frac{2n+1}{(2n\mu)^{2n/(2n+1)}} \left(1 - e^{2\phi/\sqrt{6}} \right)^{2n/(2n+1)} \right]. \quad (9.146)$$

The parameter, $\lambda = -U_{,\phi}/U$, is then given by

$$\begin{aligned} \lambda = & -\frac{4ne^{2\phi/\sqrt{6}}}{\sqrt{6}(2n\mu)^{2n/(2n+1)}} \left[1 - \frac{2n+1}{(2n\mu)^{2n/(2n+1)}} \left(1 - e^{2\phi/\sqrt{6}} \right) \right]^{-2n/(2n+1)} \\ & \times \left(1 - e^{2\phi/\sqrt{6}} \right)^{-1/(2n+1)}. \end{aligned} \quad (9.147)$$

In the deep matter-dominated epoch where the condition $R/R_c \gg 1$ is satisfied, the field ϕ is very close to zero. For n and μ of the order of unity, $|\lambda|$ is much larger than unity during this stage. Hence the matter era is realized by the instantaneous fixed point (d). As R/R_c gets smaller, $|\lambda|$ decreases to the order of unity. If the solutions reach the point $\lambda = 4Q = -4/\sqrt{6}$ and satisfy the stability condition $d\lambda/dF < 0$, then the final attractor corresponds to the de Sitter fixed point (e).

For the theories with general couplings Q , it is possible to construct a scalar-field potential that is the generalization of (9.146). One example is [273]

$$U(\phi) = U_0 [1 - C(1 - e^{-2Q\phi})^p] \quad (U_0 > 0, C > 0, 0 < p < 1). \quad (9.148)$$

Note that the $f(R)$ model (9.78) in the metric formalism corresponds to $Q = -1/\sqrt{6}$ and $p = 2n/(2n+1)$. The slope of the potential (9.148) is given by

$$\lambda = \frac{2Cp Q e^{-2Q\phi} (1 - e^{-2Q\phi})^{p-1}}{1 - C(1 - e^{-2Q\phi})^p}. \quad (9.149)$$

When $Q > 0$, the potential energy decreases from U_0 as ϕ increases from 0. On the other hand, if $Q < 0$, the potential energy decreases from U_0 as ϕ decreases from

0. In both cases we have $U(\phi) \rightarrow U_0(1 - C)$ in the limits $\phi \rightarrow \infty$ (for $Q > 0$) and $\phi \rightarrow -\infty$ (for $Q < 0$).

In the model (9.148) the field is stuck around the value $\phi = 0$ during the deep radiation and matter epochs. In these epochs one has $R \simeq \rho_m/F$ from Eqs. (9.127)–(9.129) by noting that U_0 is negligibly small compared to ρ_m or ρ_r . Using Eq. (9.129), we obtain the relation $U_{,\phi} + Q\rho_m \simeq 0$. Hence, in the high-curvature region, the field ϕ evolves along the instantaneous minima given by

$$\phi_m \simeq \frac{1}{2Q} \left(\frac{2U_0 p C}{\rho_m} \right)^{1/(1-p)}. \quad (9.150)$$

The field value $|\phi_m|$ increases for decreasing ρ_m . As long as the condition $\rho_m \gg 2U_0 p C$ is satisfied, we have $|\phi_m| \ll 1$ from Eq. (9.150).

Equation (9.149) shows that $|\lambda| \gg 1$ for field values around $\phi = 0$. Hence the instantaneous fixed point (d) can be responsible for the matter-dominated epoch provided that $|Q| \ll |\lambda|$. The variable $F = e^{-2Q\phi}$ decreases in time irrespective of the sign of the coupling Q and hence $0 < F < 1$. The de Sitter solution corresponds to $\lambda = 4Q$, i.e.

$$C = \frac{2}{(1 - F_1)^{p-1} [2 + (p - 2)F_1]}. \quad (9.151)$$

Provided that the solution of this equation exists in the region $0 < F_1 < 1$, for given values of C and p , the de Sitter solution is present.

From Eq. (9.149) we obtain

$$\frac{d\lambda}{d\phi} = - \frac{4CpQ^2 F(1 - F)^{p-2} [1 - pF - C(1 - F)^p]}{[1 - C(1 - F)^p]^2}. \quad (9.152)$$

When $0 < C < 1$ one can easily show that the function $g(F) \equiv 1 - pF - C(1 - F)^p$ is positive in the region $0 < F < 1$, giving $d\lambda/d\phi < 0$. Hence, the condition for a stable de Sitter point is automatically satisfied. In this case the solutions approach the de Sitter attractor after the end of the matter era.

When $C > 1$, the function $g(F)$ becomes negative for values of F that are smaller than the critical value $F_c (< 1)$. The de Sitter point (e) is stable under the condition $1 - pF_1 > C(1 - F_1)^p$. Using Eq. (9.151) we find that this stability condition translates to

$$F_1 > 1/(2 - p). \quad (9.153)$$

If this condition is violated, the solutions choose another stable fixed point as an attractor. In $f(R)$ theory in the metric formalism, for example, the solutions can reach the stable accelerated point (c) characterized by $m = -r - 1$ and $(\sqrt{3} - 1)/2 < m < 1$.

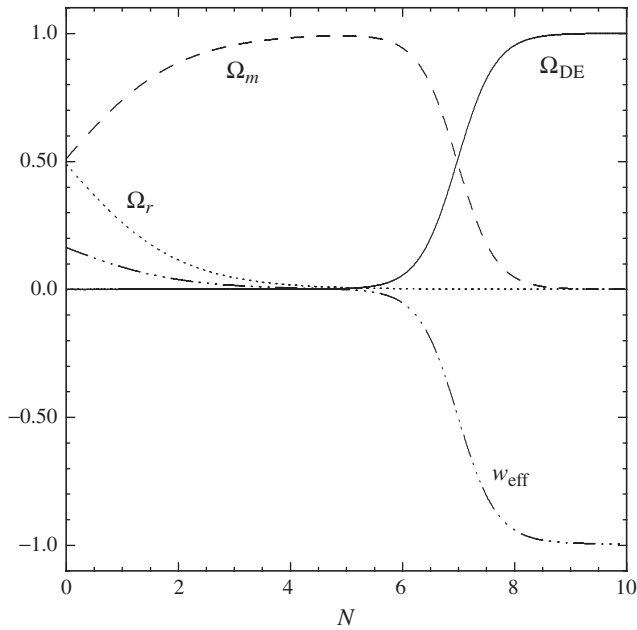


Figure 9.3 The evolution of Ω_{DE} , Ω_m , Ω_r , and w_{eff} for the field potential (9.148) with parameters $Q = 0.01$, $p = 0.2$, and $C = 0.7$. The initial conditions are given by $x_1 = 0$, $x_2 = 2.27 \times 10^{-7}$, $x_3 = 0.7$, and $x_4 - 1 = -5.0 \times 10^{-13}$ at $N = 0$. From Ref. [273].

In summary, when $0 < C < 1$, the matter point (d) can be followed by the stable de Sitter solution (e) for the model (9.148). In Fig. 9.3 we plot the evolution of Ω_{DE} , Ω_m , Ω_r , and w_{eff} versus N for $Q = 0.01$, $p = 0.2$, and $C = 0.7$. Starting from the epoch of matter–radiation equality, the solutions first dwell around the matter point (d) with $w_{\text{eff}} \simeq 0$ and finally approach the de Sitter attractor (e) with $w_{\text{eff}} \simeq -1$. Initially λ is much larger than unity and eventually approaches the value $\lambda = 4Q$.

Rewriting Eqs. (9.127) and (9.128) in the forms of Eqs. (9.44) and (9.45) by defining the dark energy density ρ_{DE} and the pressure P_{DE} to confront with SN Ia observations, one obtains the continuity equation (9.48) and the dark energy equation of state (9.49) (see problem 9.1). The variable $F = e^{-2Q\phi}$ decreases in time irrespective of the signs of Q for the model (9.148), as in the viable $f(R)$ models we discussed in Section 9.1.2. Then the phantom equation of state ($w_{\text{DE}} < -1$) as well as the cosmological constant boundary crossing ($w_{\text{DE}} = -1$) arises for scalar-tensor theories with large couplings ($|Q| = \mathcal{O}(1)$) [273]. Even if $|Q| \ll 1$ it was shown in Refs. [268, 403, 269, 264] that the phantom equation of state can be realized without introducing a ghost field with a negative kinetic energy.

9.2.2 Local gravity constraints

Let us study the local gravity constraints for the scalar-tensor theory given by the action (9.120). This is equivalent to the action (9.121) in Brans–Dicke theory with the field potential $U(\psi)$. We first consider a linear expansion of the field ψ as $\psi = \psi_0(1 + \delta_\psi)$ about a spherically symmetric background of the approximate Minkowski spacetime. In $f(R)$ theory in the metric formalism we have done a similar procedure from Eq. (9.51) to Eq. (9.61). The difference from $f(R)$ theory is that the factor “3” in Eq. (9.3) is replaced by the factor “ $3 + 2\omega_{\text{BD}}$ ” in Eq. (9.124) in Brans–Dicke theory. It is then easy to show that, in the massless limit of the field ψ (i.e. in the absence of the potential), the effective Newton constant $G_{\text{eff}}^{(\text{N})}$ and the post-Newtonian parameter γ in Brans–Dicke theory are given, respectively, by [63, 404]

$$G_{\text{eff}}^{(\text{N})} = \frac{G}{\psi_0} \frac{4 + 2\omega_{\text{BD}}}{3 + 2\omega_{\text{BD}}}, \quad \gamma = \frac{1 + \omega_{\text{BD}}}{2 + \omega_{\text{BD}}}. \quad (9.154)$$

In $f(R)$ theory in the metric formalism, Eq. (9.61) gives $G_{\text{eff}}^{(\text{N})} = (4/3)(G/\psi_0)$ and $\gamma = 1/2$ in the limit $M_\psi \rightarrow 0$, so that the Brans–Dicke parameter corresponds to $\omega_{\text{BD}} = 0$.

Using the solar system bound given in Eq. (8.171), we find that the Brans–Dicke parameter is constrained to be $\omega_{\text{BD}} > 4.3 \times 10^4$ in the massless case. This bound also applies to the case of a nearly massless field with a potential in which the Yukawa correction $e^{-M_\psi r}$ is close to 1. Using the bound $\omega_{\text{BD}} > 4.3 \times 10^4$ in Eq. (9.122), we find that the coupling Q is constrained to be

$$|Q| < 2.4 \times 10^{-3} \quad (\text{for the massless case}). \quad (9.155)$$

Under this constraint it is difficult to find a large difference from the uncoupled quintessence models.

In the presence of the field-potential it is possible for large coupling models ($|Q| = \mathcal{O}(1)$) to satisfy the local gravity constraints if the mass M of the field ϕ is sufficiently heavy in the region of high density. In fact the scalar-tensor potential (9.148) is designed to have a large mass in the high-density region so that it can be compatible with experimental tests for the violation of equivalence principle through the chameleon mechanism even for $|Q| = \mathcal{O}(1)$. In the following we shall consider the model (9.148) and derive the conditions under which the local gravity constraints can be satisfied. If we make a conformal transformation for the action (9.120), the resulting Einstein frame action is given by Eq. (8.136) with a constant coupling Q . Hence one can use the results obtained in Section 8.4.

We consider a configuration in which a spherically symmetric body has a constant density ρ_A inside the body and that the energy density outside the

body is given by $\rho = \rho_B$ ($\ll \rho_A$). Under the condition $|Q\phi| \ll 1$, one has $V_{,\phi} \simeq -2U_0 Q p C (2Q\phi)^{p-1}$ for the potential $V = U/F^2$ in the Einstein frame. Then the field values at the potential minima inside and outside the body are

$$\phi_A \simeq \frac{1}{2Q} \left(\frac{2U_0 p C}{\rho_A} \right)^{1/(1-p)}, \quad \phi_B \simeq \frac{1}{2Q} \left(\frac{2U_0 p C}{\rho_B} \right)^{1/(1-p)}, \quad (9.156)$$

which satisfy $|\phi_A| \ll |\phi_B|$. These are analogous to the field value ϕ_m derived in Eq. (9.150) in the cosmological setting. In order to realize the accelerated expansion today, U_0 needs to be roughly the same order as the square of the present Hubble parameter H_0 , so we have $U_0 \sim H_0^2 \sim \rho_c^{(0)}$, where $\rho_c^{(0)} \simeq 10^{-29} \text{ g/cm}^3$ is the present cosmological density. The baryonic/dark matter density in our galaxy corresponds to $\rho_B \simeq 10^{-24} \text{ g/cm}^3$. This shows that the conditions $|Q\phi_A| \ll 1$ and $|Q\phi_B| \ll 1$ are in fact satisfied unless $C \gg 1$. The field mass squared $m_A^2 \equiv V_{,\phi\phi}$ at $\phi = \phi_A$ is approximately given by

$$m_A^2 \simeq \frac{1-p}{(2^p p C)^{1/(1-p)}} Q^2 \left(\frac{\rho_A}{U_0} \right)^{(2-p)/(1-p)} U_0, \quad (9.157)$$

which means that m_A can be much larger than H_0 because of the condition $\rho_A \gg U_0$.

Let us place constraints on model parameters by using the bound (8.173) coming from the violation of equivalence principle. In so doing, we shall consider the case where the solutions finally approach the de Sitter point (e). The bound (8.173) coming from the violation of equivalence principle translates into

$$(2U_0 p C / \rho_B)^{1/(1-p)} < 7.4 \times 10^{-15} |Q|. \quad (9.158)$$

At the de Sitter point (e), one has $3F_1 H_1^2 = U_0[1 - C(1 - F_1)^p]$ with C given in Eq. (9.151). Hence, we obtain the following relation

$$U_0 = 3H_1^2 [2 + (p-2)F_1] / p. \quad (9.159)$$

Substituting this into Eq. (9.158) we find

$$(R_1 / \rho_B)^{1/(1-p)} (1 - F_1) < 7.4 \times 10^{-15} |Q|, \quad (9.160)$$

where $R_1 = 12H_1^2$ is the Ricci scalar at the de Sitter point. Since the term $(1 - F_1)$ is smaller than 1/2 from the condition (9.153) we obtain the inequality $(R_1 / \rho_B)^{1/(1-p)} < 1.5 \times 10^{-14} |Q|$. Taking the values $R_1 = 10^{-29} \text{ g/cm}^3$ and $\rho_B = 10^{-24} \text{ g/cm}^3$ as we have done for the $f(R)$ gravity, we get the following bound

$$p > 1 - \frac{5}{13.8 - \log_{10} |Q|}. \quad (9.161)$$

When $|Q| = 10^{-1}$ and $|Q| = 1$ we have $p > 0.66$ and $p > 0.64$, respectively. For the $f(R)$ gravity model (9.146) the above bound corresponds to $p = 2n/(2n + 1) > 0.65$, which translates into the condition $n > 0.9$ [see Eq. (9.89)].

As we will see in Section 11.6, there are cosmological constraints on the values of p and Q coming from the evolution of matter density perturbations. As long as p is close to 1, it is possible to satisfy both cosmological and local gravity constraints for $|Q| \lesssim 1$ [273].

9.3 Gauss–Bonnet dark energy models

The $f(R)$ and scalar-tensor theories add to the gravitational tensor field a new degree of freedom, a scalar field. However this certainly does not exhaust the range of possible modifications of gravity. One possibility is to add vector fields, as in e.g., Ref. [405]. Another one is to add to the Einstein Lagrangian general functions of the Ricci and Riemann tensors, e.g., $f(R, R_{\mu\nu}R^{\mu\nu}, R_{\mu\nu\alpha\beta}R^{\mu\nu\alpha\beta}, \dots)$ [406]. However these Lagrangians are generally plagued by the existence of ghosts, i.e. the existence of negative energy states [407, 408, 409]. Even beside the quantum problems, this generally implies classical instabilities either at the background or at the perturbed level.

There is however a way to modify gravity with a combination of Ricci and Riemann tensors that keeps the equations at second-order in the metric and does not necessarily give rise to instabilities, namely a Gauss–Bonnet (GB) term coupled to scalar field(s) [410, 411, 412, 413]. The GB term is a topological invariant quantity, which contributes to the dynamics in four dimensions provided that it is coupled to a dynamically evolving scalar field. It is the unique invariant for which the highest (second) derivative occurs linearly in the equations of motion, thereby ensuring the uniqueness of solutions. Note that the GB term naturally arises as a correction to the tree-level action of low-energy effective string theory [64, 210]. In what follows we shall study the possibility of realizing the late-time cosmic acceleration in the presence of the GB term coupled to a scalar field ϕ .

The model we study is given by the following action (in the unit of $\kappa^2 = 1$)

$$S = \int d^4x \sqrt{-g} \left[\frac{1}{2}R - \frac{1}{2}(\nabla\phi)^2 - V(\phi) - f(\phi)R_{\text{GB}}^2 \right] + S_m(g_{\mu\nu}, \Psi_m), \quad (9.162)$$

where $V(\phi)$ and $f(\phi)$ are functions of ϕ , and R_{GB}^2 is the GB term defined by

$$R_{\text{GB}}^2 \equiv R^2 - 4R_{\mu\nu}R^{\mu\nu} + R_{\mu\nu\alpha\beta}R^{\mu\nu\alpha\beta}. \quad (9.163)$$

The action (9.162) corresponds to the Einstein frame action in which the scalar field ϕ does not have a direct coupling to the Ricci scalar R . If this originates from

the string frame action, the field ϕ (dilaton) has a coupling with the non-relativistic matter in general. In the following we shall assume that this coupling is negligibly small (as in the case of the runaway dilaton scenario [211, 212]).

Let us study the cosmological dynamics for the action (9.162) in the flat FLRW background. We take the following exponential potential

$$V(\phi) = V_0 e^{-\lambda\phi} \quad (\lambda > 0), \quad (9.164)$$

which goes to 0 in the limit $\phi \rightarrow \infty$. The GB coupling is generally given by the sum of the exponential terms. We shall study the following simple case [410]

$$f(\phi) = (f_0/\mu) e^{\mu\phi}, \quad (9.165)$$

where f_0 and μ (> 0) are constants. As a matter action S_m we take into account non-relativistic matter and radiation, which satisfy the usual conservation equations (uncoupled to the field ϕ). The variation of the action (9.162) leads to the following equations of motion

$$3H^2 = \dot{\phi}^2/2 + V(\phi) + 24f_{,\phi}\dot{\phi}H^3 + \rho_m + \rho_r, \quad (9.166)$$

$$\ddot{\phi} + 3H\dot{\phi} - \lambda V_0 e^{-\lambda\phi} + 24f_{,\phi}H^2(H^2 + \dot{H}) = 0. \quad (9.167)$$

We define the following variables

$$x_1 \equiv \frac{\dot{\phi}}{\sqrt{6}H}, \quad x_2 \equiv \frac{\sqrt{V(\phi)}}{\sqrt{3}H}, \quad x_3 \equiv f_{,\phi}H^2, \quad x_4 \equiv \frac{\sqrt{\rho_r}}{\sqrt{3}H}, \quad (9.168)$$

together with the density parameters for each component:

$$\begin{aligned} \Omega_\phi &\equiv x_1^2 + x_2^2, & \Omega_{\text{GB}} &\equiv 8\sqrt{6}x_1x_3, & \Omega_r &\equiv x_4^2, \\ \Omega_m &\equiv 1 - \Omega_\phi - \Omega_{\text{GB}} - \Omega_r. \end{aligned} \quad (9.169)$$

Note that the term $24f_{,\phi}\dot{\phi}H^3$ in Eq. (9.166) contributes to Ω_{GB} , whereas the usual field energy density, $\rho_\phi = \dot{\phi}^2/2 + V(\phi)$, contributes to Ω_ϕ . For the functions (9.164) and (9.165) we have the following relation between x_2 and x_3 :

$$x_2^2 x_3 = \frac{f_0}{3} e^{(\mu-\lambda)\phi}. \quad (9.170)$$

Taking the time derivative of Eq. (9.166) and eliminating the term $\ddot{\phi}$ by using Eq. (9.167), we find

$$\begin{aligned} (1 - 8\sqrt{6}x_1x_3 + 96x_3^2) \frac{\dot{H}}{H^2} &= -\frac{1}{2}(3 + 3x_1^2 - 3x_2^2 + x_4^2) \\ &\quad - 4x_3(\sqrt{6}x_1 - 6\mu x_1^2 - 3\lambda x_2^2 + 24x_3). \end{aligned} \quad (9.171)$$

The variables x_i ($i = 1, 2, 3, 4$) obey the autonomous equations [414, 415, 416]

$$\frac{dx_1}{dN} = -3x_1 - 4\sqrt{6}x_3 + \frac{\sqrt{6}}{2}\lambda x_2^2 - (x_1 + 4\sqrt{6}x_3)\frac{\dot{H}}{H^2}, \quad (9.172)$$

$$\frac{dx_2}{dN} = -x_2 \left(\frac{\sqrt{6}}{2}\lambda x_1 + \frac{\dot{H}}{H^2} \right), \quad (9.173)$$

$$\frac{dx_3}{dN} = 2x_3 \left(\frac{\sqrt{6}}{2}\mu x_1 + \frac{\dot{H}}{H^2} \right), \quad (9.174)$$

$$\frac{dx_4}{dN} = -x_4 \left(2 + \frac{\dot{H}}{H^2} \right). \quad (9.175)$$

In the following we shall derive fixed points for the system (9.172)–(9.175) in the absence of radiation ($x_4 = 0$). Let us consider the cases: (i) $\lambda = \mu$ and (ii) $\lambda \neq \mu$, separately.

(i) Case: $\lambda = \mu$

When $\lambda = \mu$ one has $x_2^2 x_3 = f_0/3$ from Eq. (9.170). If $f_0 \neq 0$, neither x_2 nor x_3 is identical to zero. However, if $f_0 \ll 1$, x_2 and x_3 can be very much smaller than 1. In this case we may regard $x_2 \simeq 0$ or $x_3 \simeq 0$ as approximate fixed points for Eqs. (9.173) and (9.174). We do not consider such fixed points for the case $\lambda = \mu$, but we will discuss those points when λ is not equal to μ .

From Eqs. (9.173) and (9.174), the critical points corresponding to $x_2 \neq 0$ and $x_3 \neq 0$ obey the following equation

$$\frac{\dot{H}}{H^2} = -\frac{\sqrt{6}}{2}\lambda x_1, \quad (9.176)$$

in which case the effective equation of state defined in Eq. (4.96) is

$$w_{\text{eff}} = -1 + \frac{\sqrt{6}}{3}\lambda x_1. \quad (9.177)$$

Substituting Eq. (9.176) into Eqs. (9.171) and (9.172), the fixed points satisfy

$$\lambda(x_1^2 + x_2^2) - \sqrt{6}x_1 + 4x_3(\sqrt{6}\lambda x_1 - 2) = 0, \quad (9.178)$$

$$3(x_1^2 - x_2^2) - \sqrt{6}\lambda x_1 + 3 - 8x_1 x_3(2\sqrt{6} - 3\lambda x_1) = 0. \quad (9.179)$$

Eliminating the term x_2^2 from Eqs. (9.178) and (9.179), we find

$$(2\lambda x_1 - \sqrt{6})(8\sqrt{6}x_3 + 24\lambda x_1 x_3 - \sqrt{6}\lambda + 6x_1) = 0. \quad (9.180)$$

This gives the following two fixed points.

- (a) Scaling solution: $x_1 = \frac{\sqrt{6}}{2\lambda}$.

In this case we have $w_{\text{eff}} = 0$ from Eq. (9.177). This scaling solution is similar to the one derived in the absence of the GB coupling ($f_0 = 0$), but the field density parameter Ω_ϕ is subject to change. Substituting $x_1 = \sqrt{6}/(2\lambda)$ into Eq. (9.178), it follows that

$$x_3 = \frac{3}{16\lambda} \left[1 - \sqrt{1 - \frac{64}{27} f_0 \lambda^3} \right], \quad x_2 = \sqrt{\frac{f_0}{3x_3}}. \quad (9.181)$$

This gives

$$\Omega_\phi = \frac{3}{2\lambda^2} \left[1 + \frac{32f_0\lambda^3}{27(1 - \sqrt{1 - 64f_0\lambda^3/27})} \right], \quad \Omega_{\text{GB}} = \frac{9}{2\lambda^2} \left[1 - \sqrt{1 - \frac{64}{27} f_0 \lambda^3} \right]. \quad (9.182)$$

In the limit that $f_0 \rightarrow 0$ we get $x_2 \rightarrow \sqrt{3/(2\lambda^2)}$, $x_3 \rightarrow 0$, $\Omega_\phi \rightarrow 3/\lambda^2$, and $\Omega_{\text{GB}} \rightarrow 0$, which recover the scaling solution derived in Section 7.2.

The scaling solution may be used for the matter-dominated epoch provided that $\Omega_\phi \ll 1$, i.e. $\lambda^2 \gg 1$. In this case, however, the scaling solution is stable so that it does not exit to the phase of cosmic acceleration. If $\mu > \lambda$, the presence of the GB term can lead to the exit from the scaling matter era (as we will discuss later).

- (b) Scalar-field and GB-dominated point: $x_3 = \frac{\lambda - \sqrt{6}x_1}{8 + 4\sqrt{6}\lambda x_1}$.

We call this a “scalar-field and GB-dominated” point because the relation, $\Omega_\phi + \Omega_{\text{GB}} = 1$, holds. Using Eq. (9.178), the variable x_1 obeys the following equation

$$3(\lambda - \sqrt{6}x_1)(\sqrt{6}\lambda x_1^3 - 10x_1^2 + \sqrt{6}\lambda x_1 - 2) + 4f_0(2 + \sqrt{6}\lambda x_1)^2 = 0. \quad (9.183)$$

When $f_0 = 0$ this has a solution $x_1 = \lambda/\sqrt{6}$. It can be responsible for the late-time cosmic acceleration for $\lambda < \sqrt{2}$. Let us obtain the solution for (9.183) perturbatively under the assumption that f_0 is much smaller than unity. Substituting $x_1 = \lambda/\sqrt{6} + \epsilon$ into Eq. (9.183) under the condition $\epsilon \ll \lambda/\sqrt{6}$, we find the following approximate relation

$$x_1 \simeq \frac{\lambda}{\sqrt{6}} - \frac{4\sqrt{6}f_0(\lambda^2 + 2)}{3(6 - \lambda^2)}, \quad x_2 \simeq \sqrt{1 - \frac{\lambda^2}{6}}, \quad x_3 \simeq \frac{2f_0}{6 - \lambda^2}. \quad (9.184)$$

The effective equation of state is given by

$$w_{\text{eff}} \simeq -1 + \frac{\lambda^2}{3} - \frac{8f_0\lambda(\lambda^2 + 2)}{3(6 - \lambda^2)} \simeq -1 + \frac{\lambda^2}{3} - \frac{\lambda^2 + 2}{6} \Omega_{\text{GB}}, \quad (9.185)$$

where we have used $\Omega_{\text{GB}} \simeq 16f_0\lambda/(6 - \lambda^2)$.

In the absence of the GB coupling the late-time cosmic acceleration occurs for $\lambda^2 < 2$. In this case, however, the scaling matter era does not exist because its existence requires the condition $\lambda^2 \gg 1$. This situation does not change much even in the presence of the GB term because the condition $\lambda^2 < 6$ is required for the existence of the scalar-field

and GB-dominated point [see the variable x_2 in Eq. (9.184)]. The contribution of the GB term tends to reduce w_{eff} in Eq. (9.185), but this effect is small for $\Omega_{\text{GB}} \ll 1$.

From the above discussion the viable cosmological trajectory for the case $\lambda = \mu$ is the standard matter era ($x_i \simeq 0$ with $i = 1, 2, 3, 4$) followed by the scalar-field and GB-dominated point with $\lambda^2 \lesssim 1$ and $\Omega_{\text{GB}} \ll 1$. In this case the effect of the GB term is not important for the dynamics of dark energy.

(ii) Case: $\lambda \neq \mu$

Scaling solutions are present only for $\lambda = \mu$ in the presence of the GB coupling. Even when $\lambda \neq \mu$, however, they are regarded as approximate scaling solutions as long as the contribution of the GB term is negligibly small. This corresponds to a situation in which x_3 is very much smaller than unity. Note that x_3 cannot be exactly zero since the relation (9.170) holds. Still one can regard this as an approximate fixed point satisfying Eq. (9.170). In what follows, when we write $x_2 = 0$ or $x_3 = 0$, it means that they are not exactly zero.

- When $x_3 = 0$, we recover the five fixed points (a), (b1), (b2), (c), (d) derived in the case of the quintessence with an exponential potential in Section 7.2.
- When $x_2 = 0$ and $\sqrt{6}\mu x_1/2 = -\dot{H}/H^2$, we obtain the “kinetic and GB-dominated solution” satisfying

$$\begin{aligned} 6\mu^2 x_1^4 - (24 + \sqrt{6})\mu x_1^3 + 24\sqrt{6}(\mu^2 + 5)x_1^2 - 5\sqrt{6}\mu x_1 + 6 &= 0, \\ x_3 &= \frac{\sqrt{6}x_1(2 - \mu x_1)}{8(3\mu x_1 - \sqrt{6})}, \end{aligned} \quad (9.186)$$

with the effective equation of state:

$$w_{\text{eff}} = -1 + \frac{\sqrt{6}}{3}\mu x_1. \quad (9.187)$$

Equation (9.186) possesses two real solutions. If $\mu = 10$, for example, we get $(x_1, x_3, w_{\text{eff}}) = (8.2 \times 10^{-2}, 1.4, -3.3 \times 10^{-1})$, $(1.2 \times 10^{-1}, 2, 2 \times 10^{-2}, 1.7 \times 10^{-2})$. The cosmic acceleration ($w_{\text{eff}} < -1/3$) does not occur in either case. When $\mu > 0$ we find that the values of x_1 corresponding to two real solutions of Eq. (9.186) are larger than $\sqrt{6}/(3\mu)$, which means $w_{\text{eff}} > -1/3$ from Eq. (9.187). Hence one cannot use these solutions for dark energy.

- When $\sqrt{6}\mu x_1/2 = -\dot{H}/H^2$ and $\sqrt{6}\lambda x_1/2 = -\dot{H}/H^2$, there exists the following de Sitter fixed point

$$(x_1, x_2, x_3) = (0, 1, \lambda/8), \quad (9.188)$$

which satisfies

$$\Omega_\phi = 1, \quad \Omega_{\text{GB}} = 0. \quad (9.189)$$

The appearance of the de Sitter point comes from the presence of the GB term. Since $f_{,\phi}H^2 = \lambda/8$ and $\dot{H} = 0$ at the fixed point, the effective potential $V_{\text{eff}}(\phi)$ for the field ϕ satisfies the relation

$$\frac{dV_{\text{eff}}}{d\phi} = -\lambda V_0 e^{-\lambda\phi} + 3\lambda H_{\text{dS}}^2, \quad (9.190)$$

where H_{dS} is a Hubble parameter at the de Sitter point. The last term, which appears due to the presence of the GB term, gives rise to a potential minimum for the field ϕ .

Perturbing Eqs. (9.172)–(9.174) about the de Sitter fixed point, we obtain a 3×3 Jacobian matrix $\mathcal{M}$ for perturbations. The eigenvalues of the matrix $\mathcal{M}$ are

$$\lambda_1 = -3, \quad \lambda_{2,3} = \frac{3}{2} \left[-1 \pm \sqrt{1 + \frac{8\lambda(\lambda - \mu)}{3(2 + 3\lambda^2)}} \right]. \quad (9.191)$$

This explicitly shows the following property for the stability of the de Sitter point:

- (i) Stable for $\mu > \lambda$.
- (ii) Saddle for $\mu < \lambda$.

Provided that $\mu > \lambda$ the system approaches the stable de Sitter solution.

We are interested in the cosmological trajectories in which the scaling matter era is followed by the de Sitter solution discussed above. We require that $\mu > \lambda$ in order to exit from the scaling matter era. Let us consider the approximate scaling solution present under the conditions $x_3 \simeq 0$ and $\sqrt{6}\lambda x_1/2 = -\dot{H}/H^2$. Perturbing Eq. (9.174) about the fixed point, it follows that

$$\frac{d}{dN} \delta x_3 = \frac{3(\mu - \lambda)}{\lambda} \delta x_3. \quad (9.192)$$

If $\lambda^2 > 3$, the approximate scaling solution is stable along the perturbations of δx_1 and δx_2 (as we have seen in Section 7.2). From Eq. (9.192) the scaling matter era is a saddle for $\mu > \lambda$, whereas it is a stable node $\mu < \lambda$.

When $\mu > \lambda$ the saddle scaling solution can be followed by the stable de Sitter solution (provided that $\lambda^2 > 3$). In Fig. 9.4 we plot the evolution of Ω_ϕ , Ω_{GB} , Ω_m , and Ω_r together with w_{eff} for $\lambda = 4$, $\mu = 12$, and $f_0 = 10^{-22}$. The solution starts from a radiation-dominated epoch and connects to the approximate scaling matter era ($\Omega_m \simeq \text{constant}$). Since $\mu > \lambda$ the solution exits from the scaling regime and finally approaches the stable de Sitter fixed point. The final attractor actually satisfies $\Omega_\phi = 1$ and $\Omega_{\text{GB}} = 0$.

Figure 9.4 shows that the energy fraction of the GB term grows right after the end of the matter era, but it begins to decrease after the increase of Ω_ϕ . The effective equation of state temporally takes a local minimum value $w_{\text{eff}} \approx -0.6$ around $\Omega_m \approx 0.3$ – 0.4 , which means that the accelerated expansion occurs at the present epoch in this scenario. The rapid transition of w_{eff} just after the matter era is associated with the growth of Ω_{GB} . For increasing μ , w_{eff} tends to be smaller.

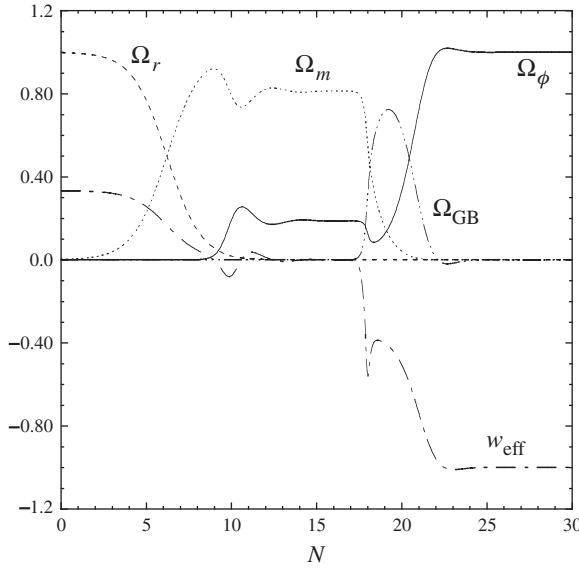


Figure 9.4 Evolution of the variables Ω_ϕ , Ω_{GB} , Ω_m , and Ω_r together with the effective equation of state w_{eff} for $\lambda = 4$, $\mu = 12$, and $f_0 = 10^{-22}$. We choose initial conditions $x_1 = 10^{-8}$, $x_2 = 10^{-7}$, $x_3 = 3.6 \times 10^{-9}$, and $x_4 = 0.999$. The solution is in a scaling regime during the matter-dominated epoch and finally approaches the de Sitter Universe characterized by $\Omega_\phi = 1$, $\Omega_{GB} = 0$, and $w_{\text{eff}} = -1$. The energy fraction of the field ϕ during the scaling regime is $\Omega_\phi = 3/\lambda^2 = 0.1875$. From Ref. [416].

When $\lambda = 4$, for example, we find that the phantom equation of state, $w_{\text{eff}} < -1$, is realized for $\mu > 25$. Meanwhile the increase of μ leads to a shorter period of the matter-dominated epoch. Hence the transient phantom stage is realized at the expense of such a short matter period. If we take larger λ , it is also difficult to get smaller values of w_{eff} satisfying the condition for the cosmic acceleration compatible with observations. Thus λ is bounded from above as well.

Koivisto and Mota [414] placed observational constraints on the above model using the Gold data set of SN Ia [109] together with the CMB shift parameter data of WMAP [14]. The parameter λ is constrained to be $3.5 \lesssim \lambda \lesssim 4.5$ at the 95% confidence level. If the solutions are in the scaling regime in the radiation era, then the constraint coming from the BBN gives $\lambda > 9.4$ under the bound $\Omega_\phi^{\text{BBN}} < 0.045$. This is relaxed down to $\lambda > 4.47$ under the milder bound $\Omega_\phi^{\text{BBN}} < 0.2$, only marginally compatible with the above constraint. In another paper [415], Koivisto and Mota included the constraints coming from the BBN, LSS, BAO, and solar system data and showed that these data strongly disfavor the GB model discussed above.

In Ref. [416] it was shown that tensor perturbations are subject to negative instabilities in the above model when the GB term dominates the cosmological dynamics. Decomposing tensor perturbations into eigenmodes of the spatial Lagrangian, $\nabla^2 e_{ij} = -k^2 e_{ij}$, with scalar amplitude $h(t)$, i.e. $h_{ij} = h(t)e_{ij}$, where e_{ij} have two polarization states, the Fourier modes of tensor perturbations obey the equation of motion [417, 418]

$$\frac{1}{a^3 Q_T} (a^3 Q_T \dot{h})' + c_T^2 \frac{k^2}{a^2} h = 0, \quad (9.193)$$

where

$$Q_T \equiv 1 - 8H\dot{f}, \quad c_T^2 \equiv \frac{1 - 8\ddot{f}}{1 - 8H\dot{f}}. \quad (9.194)$$

The stability of tensor perturbations requires the condition $c_T^2 > 0$. The no-ghost state to ensure a consistent quantum field theory demands the conditions $1 - 8\ddot{f} > 0$ and $1 - 8H\dot{f} > 0$ [407, 408] (under which the condition $c_T^2 > 0$ is ensured). For the GB model discussed above it was found in Ref. [416] that c_T^2 goes negative during the transition from the scaling matter era to the final de Sitter era. Hence the ghosts appear during the stage in which the contribution of the GB term is dominant.

The paper [419] studied local gravity constraints on the GB models with couplings of the form $f(\phi)R_{\text{GB}}^2$ and showed that the energy contribution coming from the GB term needs to be strongly suppressed to be compatible with solar-system experiments. This is typically of the order of $\Omega_{\text{GB}} \lesssim 10^{-30}$ and hence the GB term of the coupling $f(\phi)R_{\text{GB}}^2$ cannot be responsible for the current accelerated expansion of the Universe. The above discussions show that the GB term with the scalar-field coupling $f(\phi)R_{\text{GB}}^2$ can hardly be the source for dark energy.

As we mentioned, there are other classes of modified gravity models in which the Lagrangian density f is an arbitrary function of R , $P \equiv R_{\mu\nu}R^{\mu\nu}$, and $Q \equiv R_{\mu\nu\alpha\beta}R^{\mu\nu\alpha\beta}$ [406]. While it is possible to realize the cosmic acceleration with an appropriate choice of the function $f(R, P, Q)$, these theories are plagued by the appearance of spurious spin-2 ghosts unless a GB combination, i.e. $f = f(R, Q - 4P)$, is chosen [407, 408, 409]. The dark energy models based on the Lagrangian density $\mathcal{L} = R/2 + f(\mathcal{G})$, where $\mathcal{G} = R_{\text{GB}}^2$ is the GB term, have been studied by a number of authors [420, 421, 422, 423]. In order to ensure the stability of a late-time de Sitter solution and radiation/matter solutions, the most crucial condition to be satisfied is $f_{,\mathcal{G}\mathcal{G}} > 0$ [422]. This also comes from the stability of cosmological perturbations [421]. In Ref. [422] the authors presented a number of $f(\mathcal{G})$ models that are cosmologically viable at least at the background level, e.g., $f(\mathcal{G}) = \lambda(\mathcal{G}/\sqrt{\mathcal{G}_*}) \arctan(\mathcal{G}/\mathcal{G}_*) - \alpha\lambda\sqrt{\mathcal{G}_*}$, where α , λ , and $\mathcal{G}_*$ are positive

constants. These models can be consistent with solar system constraints for a wide range of parameter space [424]. However, it was found in Ref. [426] that matter perturbations in perfect fluids exhibit violent negative instabilities during radiation and matter eras. This growth of perturbations gets stronger on smaller scales, which is difficult to be compatible with the observed galaxy spectrum unless the deviation from General Relativity is extremely small. Thus $f(\mathcal{G})$ cosmological models are effectively ruled out from this Ultra-Violet instability.

9.4 Braneworld models of dark energy

In this section we shall discuss braneworld models of dark energy motivated by superstring/M-theory. In conventional Kaluza–Klein theories, extra dimensions are compactified on some manifolds in order to obtain 4-dimensional (4D) effective gravity theories. In braneworlds standard model particles are confined on a 3-dimensional brane embedded in 5-dimensional bulk spacetime with large extra dimensions. In braneworld models proposed by Randall and Sundrum [426, 427], a brane with a positive tension (σ) is embedded in a 5-dimensional Anti de Sitter (AdS) bulk. The standard 4-dimensional gravity is recovered for distances larger than the crossover scale between 4D and 5D gravity set by the AdS radius. If we consider a homogeneous and isotropic cosmology in the Randall–Sundrum model, the Friedmann equation on the brane is modified in a high-curvature regime such that the expansion law is given by $H^2 \propto \rho^2$ whereas in the low-curvature regime the standard expansion law is recovered [428, 429, 430].

There is another braneworld model proposed by Dvali, Gabadadze, and Porrati (DGP) [431]. In this approach the 3-brane is embedded in a Minkowski bulk spacetime with infinitely large extra dimensions. Newton’s law is recovered by adding a 4D Einstein–Hilbert action sourced by the brane curvature to the 5D action [432]. The presence of such a 4D term may be induced by quantum corrections coming from the bulk gravity and its coupling with matter on the brane. In the DGP model the standard 4D gravity is recovered for small distances, whereas the effect from the 5D gravity manifests itself for large distances. In particular it is possible to realize the late-time cosmic acceleration without introducing an exotic matter source [433, 434]. In this sense this solution is called the “self-accelerating Universe.” There is also a generalized version of the DGP model, see e.g., Ref. [62].

The DGP model is characterized by the following action

$$S = \frac{1}{2\kappa_{(5)}^2} \int d^5 X \sqrt{-\tilde{g}} \tilde{R} + \frac{1}{2\kappa_{(4)}^2} \int d^4 x \sqrt{-g} R - \int d^5 X \sqrt{-\tilde{g}} \mathcal{L}_m, \quad (9.195)$$

where $\tilde{g}_{AB}$ is the metric in the 5D bulk and $g_{\mu\nu} = \partial_\mu X^A \partial_\nu X^B \tilde{g}_{AB}$ is the induced metric on the brane [$X^A(x^c)$ represents the coordinates of an event on the brane

labeled by x^c], and

$$\kappa_{(5)}^2 = M_{(5)}^{-3}, \quad \kappa_{(4)}^2 = M_{(4)}^{-2}. \quad (9.196)$$

Here $M_{(5)}$ and $M_{(4)}$ are 5D and 4D Planck masses, respectively. The first and second terms in Eq. (9.195) correspond to Einstein–Hilbert actions in the 5D bulk and on the brane, respectively.

Since we are considering a Minkowski bulk, there is no contribution to the Lagrangian $\mathcal{L}_m$ from the bulk matter. Then the matter action consists of a brane-localized matter whose action is given by $\int d^4x \sqrt{-g} (\sigma + \mathcal{L}_m^{\text{brane}})$, where σ is the 3-brane tension and $\mathcal{L}_m^{\text{brane}}$ is the Lagrangian density on the brane. Since the tension is not related to the Ricci scalar R , it can be adjusted to be zero, as we do in the following.

We consider a situation in which the 3-brane is located at $y = 0$, where y is the coordinate of the 5-th dimension. In order to study cosmological dynamics on the brane, we take a metric of the form:

$$ds^2 = -n^2(\tau, y) d\tau^2 + a^2(\tau, y) \gamma_{ij} dx^i dx^j + dy^2, \quad (9.197)$$

where γ_{ij} represents a maximally symmetric spacetime with a constant curvature K . The 5D Einstein equations are given by

$$\tilde{G}_{AB} \equiv \tilde{R}_{AB} - \frac{1}{2} \tilde{R} \tilde{g}_{AB} = \kappa_{(5)}^2 \tilde{T}_{AB}, \quad (9.198)$$

where $\tilde{R}_{AB}$ is the 5D Ricci tensor, $\tilde{T}_{AB}$ is the sum of the energy-momentum tensor $T_{AB}^{(\text{brane})}$ on the brane and the contribution $\tilde{U}_{AB}$ coming from the scalar curvature of the brane:

$$\tilde{T}_{AB} = T_{AB}^{(\text{brane})} + \tilde{U}_{AB}. \quad (9.199)$$

Since we are considering a homogeneous and isotropic Universe on the brane, one can write $T_B^A{}^{(\text{brane})}$ in the form

$$T_B^A{}^{(\text{brane})} = \delta(y) \text{diag}(-\rho_M, P_M, P_M, P_M, 0). \quad (9.200)$$

Note that ρ_M and P_M are functions of τ only. The non-vanishing components coming from the Ricci scalar R of the brane are

$$\tilde{U}_{00} = -\frac{3}{\kappa_{(4)}^2} \left(\frac{\dot{a}^2}{a^2} + K \frac{n^2}{a^2} \right) \delta(y), \quad (9.201)$$

$$\tilde{U}_{ij} = -\frac{1}{\kappa_{(4)}^2} \left[\frac{a^2}{n^2} \left(-\frac{\dot{a}^2}{a^2} + 2 \frac{\dot{a}}{a} \frac{\dot{n}}{n} - 2 \frac{\ddot{a}}{a} \right) - K \right] \gamma_{ij} \delta(y), \quad (9.202)$$

where in this section a dot represents a derivative with respect to τ . For the metric (9.197) the non-vanishing components of the 5D Einstein tensor $\tilde{G}_{AB}$ are [428, 429, 433]

$$\tilde{G}_{00} = 3 \left[\frac{\dot{a}^2}{a^2} - n^2 \left(\frac{a''}{a} + \frac{a'^2}{a^2} \right) + K \frac{n^2}{a^2} \right], \quad (9.203)$$

$$\tilde{G}_{ij} = \left[a^2 \left(2 \frac{a''}{a} + \frac{n''}{n} + \frac{a'^2}{a^2} + 2 \frac{a'n'}{an} \right) + \frac{a^2}{n^2} \left(-2 \frac{\ddot{a}}{a} - \frac{a'^2}{a^2} + 2 \frac{\dot{a}\dot{n}}{an} \right) - K \right] \gamma_{ij}, \quad (9.204)$$

$$\tilde{G}_{05} = 3 \left(\frac{\dot{a}n'}{an} - \frac{\dot{a}'}{a} \right), \quad (9.205)$$

$$\tilde{G}_{55} = 3 \left(\frac{a'^2}{a^2} + \frac{a'n'}{an} \right) - \frac{3}{n^2} \left(\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2} - \frac{\dot{a}\dot{n}}{an} \right) - 3 \frac{K}{a^2}, \quad (9.206)$$

where a prime represents a derivative with respect to y .

Assuming no flow of matter along the 5-th dimension, we have $\tilde{T}_{05} = 0$ and hence $\tilde{G}_{05} = 0$. Using this relation, we find that Eqs. (9.203) and (9.206) can be written as

$$\tilde{G}_{00} = -\frac{3n^2}{2a^3a'} I', \quad \tilde{G}_{55} = -\frac{3}{2a^3\dot{a}} \dot{I}, \quad (9.207)$$

where

$$I \equiv (a'a)^2 - \frac{(\dot{a}a)^2}{n^2} - Ka^2. \quad (9.208)$$

Since we are considering the Minkowski bulk, we have $\tilde{G}_{00} = 0$ and $\tilde{G}_{55} = 0$ locally in the bulk. This then gives $I' = 0$ and $\dot{I} = 0$. The integration of these equations leads to

$$(a'a)^2 - \frac{(\dot{a}a)^2}{n^2} - Ka^2 + \mathcal{C} = 0, \quad (9.209)$$

where $\mathcal{C}$ is a constant independent of τ and y .

Let us find solutions of the Einstein equations (9.198) in the vicinity of $y = 0$. We require that the metric is continuous across the brane in order to have a well-defined geometry. Note, however, that its derivatives with respect to y can be discontinuous at $y = 0$. The Einstein tensor is made of the metric up to the second derivatives with respect to y , so the Einstein equations with a distributional source (i.e. with a Dirac's delta centered on the brane) can be written in the form [428, 429, 433]

$$g'' = T \delta(y), \quad (9.210)$$

where $\delta(y)$ is a Dirac's delta function. Integrating this equation across the brane gives

$$[g'] = T, \quad \text{where} \quad [g'] \equiv g'(0^+) - g'(0^-). \quad (9.211)$$

This shows that the jump of the first derivative of the metric is equivalent to the energy-momentum tensor on the brane.

Notice that Eqs. (9.203) and (9.204) include the second derivatives a'' and n'' of the metric. Integrating the Einstein equations, $\tilde{G}_{00} = \kappa_{(5)}^2 \tilde{T}_{00}$ and $\tilde{G}_{ij} = \kappa_{(5)}^2 \tilde{T}_{ij}$, across the brane, we obtain

$$\frac{[a']}{a_b} = -\frac{\kappa_{(5)}^2}{3} \rho_M + \frac{\kappa_{(5)}^2}{\kappa_{(4)}^2 n_b^2} \left(\frac{\dot{a}_b^2}{a_b^2} + K \frac{n_b^2}{a_b^2} \right), \quad (9.212)$$

$$\frac{[n']}{n_b} = \frac{\kappa_{(5)}^2}{3} (3P_M + 2\rho_M) - \frac{\kappa_{(5)}^2}{\kappa_{(4)}^2 n_b^2} \left(\frac{\dot{a}_b^2}{a_b^2} + 2 \frac{\dot{a}_b}{a_b} \frac{\dot{n}_b}{n_b} - 2 \frac{\ddot{a}_b}{a_b} + K \frac{n_b^2}{a_b^2} \right), \quad (9.213)$$

where the subscript “ b ” represents the quantities on the brane.

We assume the symmetry $y \leftrightarrow -y$, in which case $[a'] = 2a'(0^+)$ and $[n'] = 2n'(0^+)$. Substituting Eq. (9.212) into Eq. (9.209), we obtain the modified Friedmann equation on the brane:

$$\epsilon \sqrt{H^2 + \frac{K}{a_b^2} - \frac{\mathcal{C}}{a_b^4}} = \frac{\kappa_{(5)}^2}{2\kappa_{(4)}^2} \left(H^2 + \frac{K}{a_b^2} \right) - \frac{\kappa_{(5)}^2}{6} \rho_M, \quad (9.214)$$

where $H \equiv \dot{a}_b/(a_b n_b)$ is the Hubble parameter and $\epsilon = \pm 1$ is the sign of $[a']$. The constant $\mathcal{C}$ can be interpreted as the term coming from the 5D bulk Weyl tensor [430, 433, 434]. Since the Weyl tensor vanishes for the Minkowski bulk, we set $\mathcal{C} = 0$ in the following discussion. We also introduce a length scale

$$r_c \equiv \frac{\kappa_{(5)}^2}{2\kappa_{(4)}^2} = \frac{M_{(4)}^2}{2M_{(5)}^3}. \quad (9.215)$$

Then Eq. (9.214) can be written as

$$\frac{\epsilon}{r_c} \sqrt{H^2 + \frac{K}{a^2}} = H^2 + \frac{K}{a^2} - \frac{\kappa_{(4)}^2}{3} \rho_M, \quad (9.216)$$

where we have omitted the subscript “ b ” for the quantities at $y = 0$.

Plugging the junction conditions (9.212) and (9.213) into the (05) component of the Einstein equations, $\tilde{G}_{05} = 0$, we find that the following matter continuity equation holds on the brane (see problem 9.4):

$$\frac{d\rho_M}{dt} + 3H(\rho_M + P_M) = 0, \quad (9.217)$$

where t is the cosmic time related to the time τ via the relation $dt = n_b d\tau$. If the equation of state, $w_M = P_M/\rho_M$, is specified, the cosmological evolution is obtained by solving Eqs. (9.216) and (9.217).

For a flat geometry ($K = 0$), Eq. (9.216) reduces to

$$H^2 - \frac{\epsilon}{r_c} H = \frac{\kappa_{(4)}^2}{3} \rho_M. \quad (9.218)$$

If the Hubble radius H^{-1} is much smaller than the scale r_c (i.e. $r_c \gg H^{-1}$), the first term in Eq. (9.218) dominates over the second one. In this case the standard Friedmann equation, $H^2 = \kappa_{(4)}^2 \rho_M/3$, is recovered. Meanwhile, in the regime $r_c \lesssim H^{-1}$, the presence of the second term in Eq. (9.218) leads to a modification to the standard Friedmann equation. In the Universe dominated by non-relativistic matter ($\rho_M \propto a^{-3}$), the Universe approaches a de Sitter solution for $\epsilon = +1$:

$$H \rightarrow H_{\text{ds}} = \frac{1}{r_c}. \quad (9.219)$$

Hence it is possible to realize the present cosmic acceleration provided that r_c is of the order of the present Hubble radius H_0^{-1} . This accelerated expansion of the Universe is the result of the gravitational leakage into extra dimensions at large distances.

When $\epsilon = -1$, we do not have a late-time de Sitter solution. In the regime $r_c \ll H^{-1}$, Eq. (9.218) gives the equation $H^2 \simeq \kappa_{(5)}^2 \rho_M^2/18$. This is similar to the modified Friedmann equation which appears in the high-energy regime of the Randall–Sundrum braneworld model [426, 427]. In such cases the Universe does not exhibit the self acceleration unless a specific form of matter is introduced on the brane.

In the following let us focus on the case $\epsilon = +1$. Equation (9.216) can be written as

$$H^2 + \frac{K}{a^2} = \left(\sqrt{\frac{\kappa_{(4)}^2}{3} \rho_M + \frac{1}{4r_c^2}} + \frac{1}{2r_c} \right)^2. \quad (9.220)$$

For the matter on the brane, we consider non-relativistic matter with the energy density ρ_m and the equation of state $w_m = 0$. We then have $\rho_m = \rho_m^{(0)}(1+z)^3$ from Eq. (9.217). We introduce the following present value quantities

$$\Omega_K^{(0)} = -\frac{K}{a_0^2 H_0^2}, \quad \Omega_{r_c}^{(0)} = \frac{1}{4r_c^2 H_0^2}, \quad \Omega_m^{(0)} = \frac{\kappa_{(4)}^2 \rho_m^{(0)}}{3H_0^2}. \quad (9.221)$$

Then Eq. (9.220) reads

$$H^2(z) = H_0^2 \left[\Omega_K^{(0)}(1+z)^2 + \left\{ \sqrt{\Omega_m^{(0)}(1+z)^3 + \Omega_{r_c}^{(0)}} + \sqrt{\Omega_{r_c}^{(0)}} \right\}^2 \right]. \quad (9.222)$$

The normalization condition at $z = 0$ is given by

$$\Omega_m^{(0)} + \Omega_K^{(0)} + 2\sqrt{1 - \Omega_K^{(0)}}\sqrt{\Omega_{r_c}^{(0)}} = 1. \quad (9.223)$$

For the flat Universe ($K = 0$) this relation yields

$$\Omega_{r_c}^{(0)} = \left(\frac{1 - \Omega_m^{(0)}}{2} \right)^2. \quad (9.224)$$

If one introduces an effective dark energy component with $\rho_{\text{DE}} \equiv 3H/(\kappa_{(4)}^2 r_c)$, one can use the continuity equation $\dot{\rho}_{\text{DE}} + 3H(1 + w_{\text{DE}})\rho_{\text{DE}} = 0$ to define an effective equation of state for $K = 0$ [437]

$$w_{\text{DE}}(z) = \frac{\Omega_m^{(0)} - 1 - \sqrt{(1 - \Omega_m^{(0)})^2 + 4\Omega_m^{(0)}(1+z)^3}}{2\sqrt{(1 - \Omega_m^{(0)})^2 + 4\Omega_m^{(0)}(1+z)^3}}. \quad (9.225)$$

For small redshifts one can approximate $w_{\text{DE}}(z) = w_0 + w_a z/(1+z)$ with $w_0 \approx -(1 + \Omega_m^{(0)})^{-1}$ and $w_a \approx 3\Omega_m^{(0)}(1 - \Omega_m^{(0)})/(1 + \Omega_m^{(0)})^3$. For instance for $\Omega_m^{(0)} = 0.3$ we have $w_0 \approx -0.77$ and $w_a \approx 0.29$.

The parametrization (9.222) of the Hubble parameter together with the normalization (9.223) can be used to place observational constraints on the DGP model at the background level [435, 436, 437, 439, 438]. In Fig. 9.5 we show the joint constraints from observational data of SNLS, BAO, and the CMB shift parameter [437]. While the flat DGP model can be consistent with the SN Ia data, it is under the strong observational pressure by adding the data of the BAO and the CMB shift parameter. The open DGP model gives a slightly better fit relative to the flat model [437, 439].

We have to caution that the above constraints have been derived by considering the cosmic expansion history at the background level. In order to obtain precise constraints coming from the power spectrum of BAO and LSS, we need to know the evolution of density perturbations in the DGP model. As we will see in Section 11.6, a quasi-static approximation to the 5D cosmological perturbations shows that the DGP model contains a ghost mode in the scalar sector of the gravitational field [440, 441, 442, 443, 444]. Thus the original DGP model is effectively ruled out as a viable dark energy model by the observational pressure and by the ghost problem. It is however possible to construct a generalized version of the DGP model free

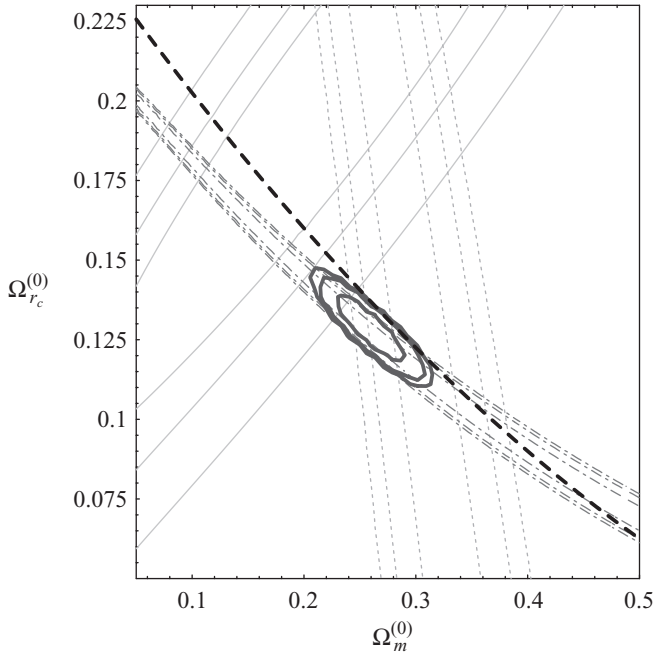


Figure 9.5 Combined observational constraints on the DGP model from the SNLS data (solid thin), the BAO (dotted), and the CMB shift parameter from the WMAP 3-year data (dot-dashed). The thick line represents the curve (9.224) for the flat model ($\Omega_K^{(0)} = 0$). From Ref. [437].

from the ghost problem by embedding our visible 3-brane with a 4-brane in a flat 6D bulk [176].

9.5 Problems

9.1 Let us consider the following general action

$$S = \int d^4x \sqrt{-g} \left[\frac{1}{2} f(R, \phi, X) + \mathcal{L}_m \right], \quad (9.226)$$

where f is a function in terms of a Ricci scalar R , a scalar field ϕ , and a kinetic term $X = -(1/2)(\nabla\phi)^2$. For the matter Lagrangian $\mathcal{L}_m$, we take into account non-relativistic matter with an energy density ρ_m . Derive field equations in the flat FLRW background. Find an energy density ρ_{DE} and a pressure P_{DE} to confront with SN Ia observations and show that these satisfy the continuity equation $\dot{\rho}_{DE} + 3H(\rho_{DE} + P_{DE}) = 0$.

9.2 From the trace equation (9.124) in the Brans–Dicke theory, derive an equation for the linear perturbation δ_ψ and the field mass squared M_ψ^2 by decomposing the field ψ into the background and perturbed parts, $\psi = \psi_0(1 + \delta_\psi)$, in the Minkowski background.

The $f(R)$ theory in the metric formalism corresponds to the Brans–Dicke theory with $\omega_{\text{BD}} = 0$, $\psi = F(R)$, and $U = (1/2)(RF - f)$. Show that the mass squared derived above agrees with Eq. (9.52) by using this correspondence.

- 9.3** Find the eigenvalues of the 2×2 matrix of perturbations about the fixed points (a)–(e) for the scalar-tensor action (9.120) with $F(\phi) = e^{-2Q\phi}$.
- 9.4** Derive Eq. (9.217) by substituting Eqs. (9.212) and (9.213) into the (05) component of the Einstein equations, $\tilde{G}_{05} = 0$.

Cosmic acceleration without dark energy

The crucial kick to dark energy research was the interpretation in 1998 of standard candle observations in terms of cosmic acceleration in the FLRW metric. What we observe is however merely that distant sources ($z > 0.3$) recede slower than we would predict in an Einstein–de Sitter Universe calibrated through “nearby” sources. That is, we observe different expansion rates at different distances rather than an increase in the expansion rate at all distances. Can this be caused by a strong inhomogeneity rather than by an accelerating Universe?

We also noticed that cosmic acceleration seems to be a recent phenomenon, at least for standard dark energy models. The epoch in which dark energy begins to play a role is close to the epoch in which most of the cosmic structures formed out of the slow linear gravitational growth. We are led to ask again: can the acceleration be caused by strong inhomogeneities rather than by a dark energy component?

The answer to both questions is yes, at least in principle. First, we can always interpret a homogeneous evolution $H(z)$ as a line-of-sight inhomogeneous rate $H(r)$ since we observe only along our past light cone $ds^2 = 0$ and time and distance are inextricably related. Second, one can always arrange matter sources so that in some region of the Universe they accelerate away from each other even if on larger scales the expansion is decelerated. In both cases, the price to pay is to allow for huge inhomogeneities while still retaining a compatible level of isotropy. Although no model that could convincingly explain either has been proposed so far, the motivation for linking inhomogeneities with dark energy is so appealing that some of the ideas introduced in this context are worth exploring. Among many blind alleys we could perhaps encounter some good escape route.

In this section primes and dots refer to partial space and time derivatives, respectively.

10.1 Void models

The basic motivation for a void model was presented soon after the supernovae discovery [445, 446, 447, 448]. Since we observe that nearby SN Ia recede faster than the $H(z)$ predicted by the Einstein–de Sitter Universe, we could assume that we live in the middle of a huge spherical region which is expanding faster because it is emptier than the outside. The transition redshift z_e , i.e. the void edge, should be located around 0.3–0.5, the value at which in the standard interpretation we observe the beginning of acceleration.

The consistent way to realize such a spherical inhomogeneity has been studied since the 1930s in the relativistic literature: the Lemaître–Tolman–Bondi (LTB) metric. This is the generalization of a FLRW metric in which the expansion factor along the radial coordinate r is different relative to the surface line element $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$. If we assume the inhomogeneous metric

$$ds^2 = -dt^2 + X^2(t, r) dr^2 + R^2(t, r) d\Omega^2, \quad (10.1)$$

and solve the (0, 1) Einstein equation for a fluid at rest (see problem 10.1), we find that the LTB metric is given by

$$ds^2 = -dt^2 + \frac{[R'(t, r)]^2}{1 + \beta(r)} dr^2 + R^2(t, r) d\Omega^2, \quad (10.2)$$

where $R(t, r)$, $\beta(r)$ are arbitrary functions. We remind the reader again that in this section primes and dots refer to partial space and time derivatives, respectively. The function $\beta(r)$ can be thought of as a position-dependent spatial curvature. If R is factorized so that $R(t, r) = a(t)f(r)$ and $\beta(r) = -Kf^2(r)$, then we recover the FLRW metric (up to a redefinition of r : from now on when we seek the FLRW limit we put $R = a(t)r$ and $\beta = -Kr^2$). Otherwise, we have a metric representing a spherical inhomogeneity centered on the origin. An observer located at the origin will observe an isotropic Universe. We can always redefine r at the present time to be $R_0 \equiv R(t_0, r) = r$, so that the metric is very similar to a FLRW today.

Considering the infinitesimal radial proper length $D_{||} = R' dr / \sqrt{1 + \beta}$, we can define the *radial Hubble function* as

$$H_{||} \equiv \dot{D}_{||} / D_{||} = \dot{R}' / R', \quad (10.3)$$

and similarly the *transverse Hubble function*:

$$H_{\perp} = \dot{R} / R. \quad (10.4)$$

Of course the two definitions coincide for the FLRW metric. The non-vanishing components of the Ricci tensor for the LTB metric are

$$R_0^0 = \frac{2\ddot{R}}{R} + \frac{\ddot{R}'}{R'}, \quad (10.5)$$

$$R_1^1 = \frac{2\dot{R}\ddot{R}' + R\ddot{R}' - \beta'}{RR'}, \quad (10.6)$$

$$R_2^2 = R_3^3 = \frac{\dot{R}^2 - \beta}{R^2} + \frac{\dot{R}\ddot{R}' + R'\ddot{R} - \beta'/2}{RR'}. \quad (10.7)$$

In terms of the two Hubble functions, we find that the Friedmann equations for the pressureless matter density $\rho_m(t, r)$ are given by [449]

$$H_\perp^2 + 2H_\parallel H_\perp - \frac{\beta}{R^2} - \frac{\beta'}{RR'} = 8\pi G\rho_m, \quad (10.8)$$

$$6\frac{\ddot{R}}{R} + 2H_\perp^2 - 2\frac{\beta}{R^2} - 2H_\parallel H_\perp + \frac{\beta'}{RR'} = -8\pi G\rho_m. \quad (10.9)$$

Adding Eqs. (10.8) and (10.9), it follows that $2R\ddot{R} + \dot{R}^2 = \beta$. Integrating this equation, we obtain a Friedmann-like equation

$$H_\perp^2 = \frac{\alpha(r)}{R^3} + \frac{\beta(r)}{R^2}, \quad (10.10)$$

where $\alpha(r)$ is a free function that we can use along with $\beta(r)$ to describe the inhomogeneity. From this we can define an effective density parameter $\Omega_m^{(0)}(r) = \Omega_m(r, t_0)$ today:

$$\Omega_m^{(0)}(r) \equiv \frac{\alpha(r)}{R_0^3 H_{\perp 0}^2}, \quad (10.11)$$

where $H_{\perp 0} \equiv H_\perp(r, t_0)$, and the effective spatial curvature is given by

$$\Omega_K^{(0)}(r) = 1 - \Omega_m^{(0)}(r) = \frac{\beta(r)}{R_0^2 H_{\perp 0}^2}. \quad (10.12)$$

Hence we see that the initial condition at some time t_0 (which here we take as the present time) must specify two free functions of r , for instance $\alpha(r)$, $\beta(r)$ or $\Omega_m^{(0)}(r)$, $H_{\perp 0}(r)$. The latter choice shows that the inhomogeneity can be in the matter distribution or in the expansion rate or in both. This freedom can be used to fit simultaneously for any expansion rate (and therefore luminosity and angular diameter distances [450]) and for any source number density [451].

If one imposes the additional constraint that the age of the Universe is the same for every observer (see problem 10.2), then only one free function is left [452]. The same occurs if one chooses $\Omega_m^{(0)}(r) = \text{constant}$ (notice that this is different from $\rho_m^{(0)}(r) = \text{constant}$, which is another possible choice) i.e. if the matter density

fraction is assumed homogeneous today (and only today) [453]. The choice of a homogeneous Universe age guarantees against the existence of diverging inhomogeneities in the past. However, there is no compelling reason to impose such restrictions.

Equation (10.10) is the classical cycloid equation whose solution for $\beta > 0$ is given parametrically by

$$R(r, \eta) = \frac{\alpha(r)}{2\beta(r)}(\cosh \eta - 1) = \frac{R_0 \Omega_m^{(0)}(r)}{2[1 - \Omega_m^{(0)}(r)]}(\cosh \eta - 1), \quad (10.13)$$

$$t(r, \eta) - t_B(r) = \frac{\alpha(r)}{2\beta^{3/2}(r)}(\sinh \eta - \eta) = \frac{\Omega_m^{(0)}(r)}{2[1 - \Omega_m^{(0)}(r)]^{3/2} H_{\perp 0}}(\sinh \eta - \eta), \quad (10.14)$$

where $t_B(r) = t(r, \eta = 0)$ is the inhomogeneous “big bang” time, i.e. the time for which $\eta = 0$ and $R = 0$ for a point at comoving distance r . This can be put to zero in all generality by a redefinition of time. The “time” variable η is defined by the relation

$$\eta = \int^t \frac{\beta(r)^{1/2}}{R(\tilde{t}, r)} d\tilde{t}. \quad (10.15)$$

Notice that the “time” η that corresponds to a given t depends on r ; so $R(r, t)$ is found by solving numerically $\eta(t, r)$ from Eq. (10.14) and then substituting $R[r, \eta(r, t)]$. The present epoch $\eta_0(r)$ is defined by the condition $R = R_0$. In problem 10.2 we will derive the age of the Universe $t_{\text{age}}(r) = t(r, \eta_0) - t_B(r)$ in terms of $\Omega_m^{(0)}$, $H_{\perp 0}$. For $\beta < 0$ the η functions in Eqs. (10.13) and (10.14) become $(1 - \cos \eta)$ and $(\eta - \sin \eta)$ for R and t , respectively, while for $\beta = 0$ they are $\eta^2/2$ and $\eta^3/6$: we will not consider these cases further.

Since we need to have a faster expansion inside some distance to mimic cosmic acceleration, we need to impose to our solution the structure of a void. An example of the choice of $\Omega_m^{(0)}(r) \equiv \Omega_m(r, t_0)$, $h^{(0)}(r) \equiv H_{\perp 0}/(100 \text{ km sec}^{-1} \text{ Mpc}^{-1})$ is [452]

$$\Omega_m^{(0)}(r) = \Omega_{\text{out}} + (\Omega_{\text{in}} - \Omega_{\text{out}})f(r, r_0, \Delta), \quad (10.16)$$

$$h^{(0)}(r) = h_{\text{out}} + (h_{\text{in}} - h_{\text{out}})f(r, r_0, \Delta), \quad (10.17)$$

with

$$f(r, r_0, \Delta) = \frac{1 - \tanh[(r - r_0)/2\Delta]}{1 + \tanh(r_0/2\Delta)}, \quad (10.18)$$

representing the transition function of a shell of radius r_0 and thickness Δ . The six constants Ω_{in} , Ω_{out} , h_{in} , h_{out} , r_0 , Δ completely fix the model. If $h_{\text{in}} > h_{\text{out}}$ we can mimic the accelerated expansion.

In order to compare the LTB model to observations we need to generalize two familiar concepts: redshift and luminosity distance. The propagation of photons is described by the geodesic equations (4.213)–(4.214) discussed in Section 4.11. In the LTB metric, the geodesic equations for an incoming photon can be written as [455]

$$\frac{dt}{d\lambda_s} = -\sqrt{\frac{(R')^2}{1+\beta}} p^2 + \frac{J^2}{R^2}, \quad (10.19)$$

$$\frac{d\theta}{d\lambda_s} = \frac{J}{R^2}, \quad (10.20)$$

$$\frac{dp}{d\lambda_s} = 2\dot{R}' p \sqrt{\frac{p^2}{1+\beta} + \frac{J^2}{R^2 R'^2}} + \frac{1+\beta}{R^3 R'} J^2 + \left(\frac{\beta'}{2+2\beta} - \frac{R''}{R'} \right) p^2, \quad (10.21)$$

where λ_s is the affine parameter, $p \equiv dr/d\lambda_s$, and

$$J \equiv R^2 \frac{d\theta}{d\lambda_s} = \text{constant} = J_0 \quad (10.22)$$

is the conserved angular momentum that vanishes for a radial propagation (which is the case only for the observer at the center as we are assuming now). For $J = 0$ the photon time-distance law $t_p(r)$ can be found by dividing $dt/d\lambda_s$ by $dr/d\lambda_s$, so we obtain

$$\frac{dt_p}{dr} = -\frac{R'}{\sqrt{1+\beta}}. \quad (10.23)$$

We impose the condition $t_p(r=0) = t_0$ at the epoch of the observation. In problem 10.3 you will find that $z(r)$ is given by the solution to the following equation

$$\frac{dz}{dr} = (1+z) \frac{\dot{R}'}{\sqrt{1+\beta}}, \quad (10.24)$$

where $R(t, r)$ must be calculated on the trajectory $t_p(r)$ and we must impose $z(r=0) = 0$. Every LTB function, e.g., $H_\perp(t, r)$, $R(t, r)$ etc., can be converted into line-of-sight functions of redshift by evaluating the arguments $r_p(z)$, $t_p(z)$ along the past light cone.

The proper area of an infinitesimal surface at $r, t = \text{constant}$ is given by $A = R^2(r, t) \sin \theta d\theta d\phi$. The angular diameter distance is the square root of $A/(\sin \theta d\theta d\phi)$ so that $d_A(z) = R(t_p(z), r_p(z))$. Since the duality relation $d_L = (1+z)^2 d_A$ remains valid in inhomogeneous models, we have [454]

$$d_L(z) = (1+z)^2 R(t_p(z), r_p(z)). \quad (10.25)$$

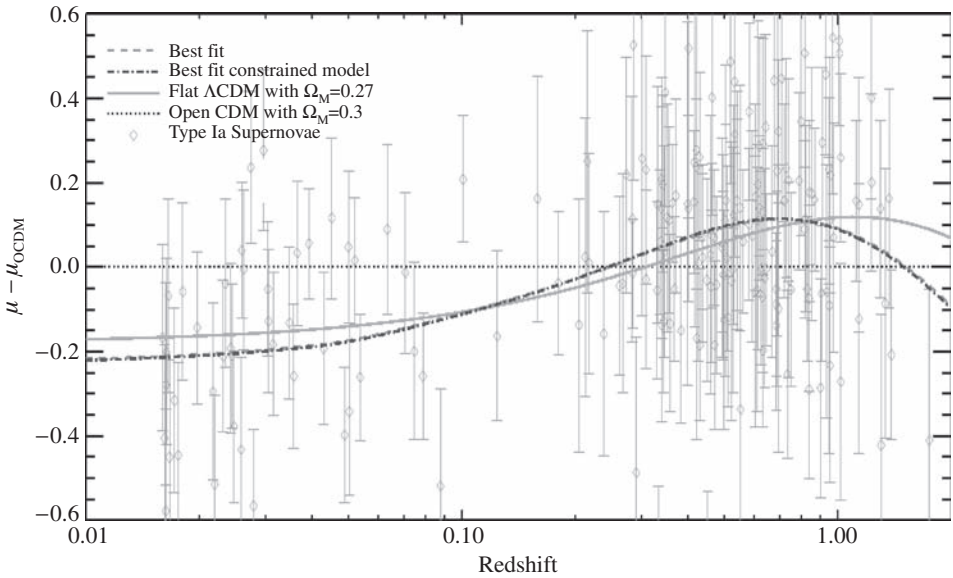


Figure 10.1 Apparent magnitude residuals for two best fit LTB models (with and without the constraint of a uniform big bang time), the best fit Λ CDM FLRW model, and the standard open CDM model, compared to Type Ia Supernovae data. From Ref. [452].

This clearly reduces to $d_L = (1+z)r(z)$ in the FLRW background. Armed with these observational tools, we can compare any LTB model to the observations.

Beside matching the SN Ia Hubble diagram, we do not want to spoil the CMB acoustic peaks and we also need to impose a local density Ω_{in} near 0.1–0.3, a flat space outside (to fulfill inflationary predictions), i.e. $\Omega_{\text{out}} = 1$, and finally the observed local Hubble value $h_{\text{in}} \approx 0.7 \pm 0.1$. The CMB requirement can be satisfied by a small value of h_{out} , since we know that to compensate for $\Omega_{\text{out}} = 1$ we need a small Hubble rate (remember that the CMB essentially constrains $\Omega_m^{(0)}h^2$). This fixes $h_{\text{out}} \approx 0.5$. So we are left with only r_0 and Δ to be constrained by SN Ia. As anticipated we expect r_0 to be near $z = 0.5$, which in the standard Λ CDM model gives a distance $r(z) \approx 2$ Gpc. An analysis using SN Ia data [452] finds that $r_0 = 2.3 \pm 0.9$ Gpc and $\Delta/r_0 > 0.2$, see Fig. 10.1. Interestingly, a “cold spot” in the CMB sky could be attributed to a void of comparable size [456, 457].

There are many more constraints one can put on such large inhomogeneities. Matter inside the void moves with respect to CMB photons coming from outside. So the hot intracluster gas will scatter the CMB photons with a large peculiar velocity and this will induce a strong kinematic Sunyaev–Zel’dovich effect [458].

Moreover, secondary photons scattered towards us by reionized matter inside the void should also distort the black body spectrum due to the fact that the CMB radiation seen from anywhere in the void (except from the center) is anisotropic and therefore at different temperatures [459]. These two constraints require the voids not to exceed 1 or 2 Gpc, depending on the exact modeling, and they are therefore already in mild conflict with the fit to supernovae.

Other constraints will follow soon with more data and more analyses. For instance, the whole process of structure formation should be revisited in an LTB Universe and this will certainly add more constraints. On the other hand, the void models can be rendered more realistic by gathering a Swiss-cheese collection of many LTB structures [460, 461].

Moreover, while in the FLRW background the function $H(z)$ fixes the comoving distance $\chi(z)$ up to a constant curvature (and consequently also the luminosity and angular diameter distances), in the LTB model the relation between $\chi(z)$ and $H_\perp(z)$ or $H_\parallel(z)$ can be arbitrary. That is, one can choose the two spatial free functions to be for instance $H_\perp(r, 0)$ and $R(r, 0)$, from which the line-of-sight values $H_\perp(z)$ and $\chi(z)$ would also be arbitrarily fixed. This shows that the “consistency” FLRW relation between $\chi(z)$ and $H(z)$ is violated in the LTB model, and in general in any strongly inhomogeneous Universe.

To see this we need now to use again the comoving angular diameter distance defined in Eq. (5.30), $d_A^{(c)} = (1+z)d_A$, but for simplicity of notation we just denote it as d . The duality relation $d_L = (1+z)^2 d_A$ leads to $d = d_L/(1+z)$ and hence

$$d(z) = \frac{1}{H_0 \sqrt{\Omega_K^{(0)}}} \sinh \left(\sqrt{\Omega_K^{(0)}} \int_0^z \frac{d\tilde{z}}{E(\tilde{z})} \right), \quad (10.26)$$

where we have used the luminosity distance (2.68) in the FLRW background. By differentiation of Eq. (10.26) the following consistency relation valid at any redshift can be derived [462]:

$$\Omega_K^{(0)} = \frac{[H(z)d_{,z}(z)]^2 - 1}{[H_0 d(z)]^2}, \quad (10.27)$$

where “ $_{,z}$ ” represents a derivative with respect to z . Differentiating this equation again we obtain the following relation in the FLRW background:

$$1 + H^2 (dd_{,zz} - d_{,z}^2) + HH_{,z} dd_{,z} = 0. \quad (10.28)$$

Since both $H(z)$ and $d(z)$ are observable, for instance with the baryon acoustic oscillations, this test can be employed to detect large-scale deviations from homogeneity. The LTB model violates this relation or, equivalently, shows an apparent present curvature $\Omega_K^{(0)}$ that depends on the redshift. By differentiating Eq. (10.25)

and using Eqs. (10.23) and (10.24), one finds for $d = d_L/(1+z) = (1+z)R$:

$$\begin{aligned} d_{,z} &= \frac{d}{1+z} + \frac{d}{R} \left(\dot{R} \frac{dt}{dz} + R' \frac{dr}{dz} \right) \\ &= \frac{d}{1+z} \left(1 - \frac{H_{\perp}}{H_{\parallel}} \right) + \frac{\sqrt{1+\beta}}{H_{\parallel}}. \end{aligned} \quad (10.29)$$

Since from Eq. (10.12) we have $\beta = R_0^2 H_{\perp 0}^2 \Omega_K^{(0)}$, we obtain

$$\Omega_K^{(0)} = \frac{H_{\parallel}^2}{H_{\perp 0}^2} \frac{d_{,z}^2}{R_0^2} \left[1 - \frac{\Delta H}{H_{\parallel}} \frac{d \log(1+z)}{d \log d} \right]^2 - \frac{1}{H_{\perp 0}^2 R_0^2}, \quad (10.30)$$

where $\Delta H = H_{\parallel} - H_{\perp}$. In FLRW $R_0 = d$ and $\Delta H = 0$ and hence we recover Eq. (10.27).

Finally, a radial inhomogeneity can in principle be distinguished from a homogeneous Universe by the redshift drift, an effect we will study more in detail in Section 14.6.4. The redshift drift in fact probes the local expansion rate and therefore separates effects due to variations along the time direction from those along spatial hypersurfaces [463].

Centered observers violate the Copernican principle of “non-special” position to the maximal extent. Off-center observers, however, have their own problems. They should see direct evidence of anisotropy, either in the distribution of distant sources or in the Hubble diagram of supernovae or as an additional CMB dipole [455] or even as an apparent parallax of quasars [464]. If the observed CMB dipole were entirely due to the LTB anisotropy of an off-center observer, the maximum distance to the center would be severely constrained, $r_c < 10\text{--}20$ Mpc, although one cannot exclude the possibility of a chance cancellation with the Sun’s own motion.

Finally, we should not forget that there is no valid mechanism at present to explain the formation of such huge inhomogeneities, let alone one with our Galaxy near the center. More than anything else, void models are useful to remind us how many options are still open – and at the same time how difficult it is to find one that really works.

10.2 Backreaction

Void models aim to explain the supernovae data with an *apparent* acceleration induced by a strong inhomogeneity. There is however a line of research that tries to explain cosmic acceleration by arranging inhomogeneities so that the deviation from the FLRW metric can produce a *real* acceleration, at least in some region. The whole idea rests on two facts: that the GR equations are non-linear, and that our real world is far from homogeneous, at least on small scales and perhaps

on super-horizon scales. Because of this, averaging the inhomogeneities and then solving the GR equations (the usual approach) might not be the same as first solving the full inhomogeneous GR equations and then averaging them. In other words, the expected value of a non-linear function of x is not the same as the non-linear function of the expected value of x . This effect is often called *backreaction* because one looks for the effect of inhomogeneities on background expansion [465, 466], although properly speaking there is no *reaction* in the first place since inhomogeneities are not caused by the expansion itself.

The argument is very complex, still in full evolution and has caused bitter controversy; as such it is more suited to a review than to an introductory book and we refer the reader to several published accounts [465, 466, 467, 468, 469, 470, 471, 472, 473, 474]. The underlying ideas are however rather straightforward and could serve as inspiration to the interested reader to look further.

Let us start with the Einstein equations

$$G_{\mu\nu} = 8\pi G T_{\mu\nu}. \quad (10.31)$$

We can expand both sides at first order:

$$G_{\mu\nu} = G_{\mu\nu}^{(0)} + G_{\mu\nu}^{(1)}, \quad T_{\mu\nu} = T_{\mu\nu}^{(0)} + T_{\mu\nu}^{(1)}, \quad (10.32)$$

where we take the FLRW metric as zero-th-order Universe. For pressureless matter, the (0,0) equation can be written as

$$G_{00}^{(0)} = 8\pi G (T_{00}^{(0)} + T_{00}^{(1)}) - G_{00}^{(1)}. \quad (10.33)$$

Now, if we observe some average matter density $\langle\rho\rangle$, we can identify at this order

$$\langle\rho\rangle = T_{00}^{(0)} + T_{00}^{(1)}, \quad (10.34)$$

and by averaging the full equation we obtain

$$\langle G_{00}^{(0)} \rangle = 8\pi G \langle\rho\rangle - \langle G_{00}^{(1)} \rangle, \quad (10.35)$$

where $G_{00} = 3H^2$ is the usual quantity in the flat FLRW background. We immediately see that $3H^2 \neq 8\pi G \langle\rho\rangle$, as we are accustomed to see. The reason is that normally we first average the metric, thereby obtaining $\langle g_{\mu\nu}^{(1)} \rangle = 0$ and then calculate $G_{\mu\nu}$. This gives obviously $G_{00}^{(1)} = G_{00}(\langle g_{\mu\nu}^{(1)} \rangle) = 0$. Similarly, at second order we could write

$$\langle G_{00}^{(0)} \rangle = 8\pi G \langle\rho\rangle - \langle G_{00}^{(1)} + G_{00}^{(2)} \rangle, \quad (10.36)$$

and so on.

So, where is all this leading us? If the additional terms induce an acceleration on the background expansion, then we would have a direct causal link between the rise of perturbations and the observed acceleration and, of course, we would not

require a dark energy field. This idea is extremely appealing, but are we any closer to its realization?

There are three problems here, simple to express but extremely hard to answer (which is a good sign for a scientist). They are: (a) how do we take an average?, (b) at which perturbative order should we stop?, and (c) why, if any order is so important as to change the background expansion, don't we see such huge inhomogeneities?

The average problem is the basic one. Ref. [466] suggested a reasonable-looking average of a function $f(t, x_i)$:

$$\langle f \rangle(t) = \frac{\int d^3x \sqrt{\gamma(t, x^i)} f(t, x^i)}{\int d^3x \sqrt{\gamma(t, x^i)}}, \quad (10.37)$$

where γ is the determinant of the perturbed metric of the spatial constant- t hypersurfaces. By using this average one obtains at second order a number of terms in $\langle G_{00}^{(2)} \rangle$ that could contribute to the expansion rate, because of long or small wavelengths. Typically the contribution is small, of order 10^{-5} , by assuming standard power spectra. However, this is much larger than the naive expectation of something of the order of $\delta^2 \approx 10^{-10}$ based on the fluctuation at Hubble scales and, moreover, could be enhanced by adding more and more higher-order terms or by unknown super-Hubble sized fluctuations. Nevertheless, the average is performed on constant-time hypersurfaces and not on the light cone and this is suspicious by itself, since it implies that an instantaneous average can affect the whole cosmic expansion. Changing the averaging procedure alters the result and the effect may disappear or change in amplitude [470].

Then of course, if anything appears at second order that is of any importance, one should investigate higher-order terms, as in any perturbative approach. Ref. [468] has shown that indeed many contributions cancel each other and has discussed in specific cases the danger of arbitrarily stopping at some order.

Finally, every model based on large inhomogeneities must provide a way to conceal them from our sight, at least to some extent. This can be done in part by assuming strong peculiar velocities instead of strong density fluctuations. We have learned in the void model that this is always possible, at least at one given epoch. As we have seen, however, there are also strong constraints on peculiar velocities from e.g., the kinematic Sunyaev–Zel'dovich effect. Moreover, the accompanying anisotropy is another source of observable effects difficult to accommodate with current observations.

All these issues leave the effective impact of backreaction in a very uncertain status. As we mentioned at the beginning of this section, this area of research is very active, very interesting, and very controversial. It is likely it will remain so for many years.

In addition to the models we have discussed in this chapter, there is another class of apparent acceleration models without dark energy based on the “Ultra Strong” version of the equivalence principle [475, 476]. In this scenario the usual geometric description of spacetime as a metric manifold is just a small distance approximation – on large scales General Relativity is modified by a curvature-dependent subleading effect such that the luminosity distance increases. It will be of interest to see whether such Infra-Red modifications of gravity can be consistent with SN Ia and other observations without introducing a dark energy component.

10.3 Problems

- 10.1** Find the g_{rr} component for the LTB metric by solving $R_{01} = 0$.
- 10.2** Derive the age of the Universe in the LTB model. Under which condition is it the same for every observer?
- 10.3** Let us consider an electromagnetic wave emitted during time interval $\varepsilon(0)$ and observed during a time interval $\varepsilon(\lambda_s)$, where λ_s is the affine parameter along the geodesic. The function $\varepsilon(\lambda_s)$ is therefore the difference between two solutions $t_1(\lambda_s)$ and $t_2(\lambda_s) = t_1(\lambda_s) + \varepsilon(\lambda_s)$ along the same trajectory of the geodesic equations. The redshift is then

$$z \equiv \frac{\varepsilon(0) - \varepsilon(\lambda_s)}{\varepsilon(\lambda_s)} . \quad (10.38)$$

From this expression, find the redshift equation (10.24).

11

Dark energy and linear cosmological perturbations

Most of the previous chapters explored general properties of dark energy models that are connected to their background behavior. However, dark energy influences not only the expansion rate of the Universe but also the growth of perturbations, so to this we turn now our attention.

In this chapter we discuss several advanced topics about linear cosmological perturbations. These include (i) perturbations for a general dark energy fluid, (ii) perturbations for a dark energy scalar field, and (iii) perturbations in modified gravity models.

Throughout this chapter, a prime represents a derivative with respect to $N = \ln a$ (not to conformal time as in previous chapters), unless otherwise specified.

11.1 Perturbations in a general dark energy cosmology

The linear perturbation equations we have derived in Chapter 4 for a single fluid and for two-fluid cases can be generalized in several ways, such as considering more fluids, interaction terms, and various levels of approximations, but the physics and the mathematics involved are more or less always the same. The present Universe is well described by at least two components, matter and dark energy, where the latter is completely unknown. It is then useful to derive the equations in a very general case by assuming a general equation of state $w(z)$ and a general sound speed $c_s^2(z)$ in a multi-fluid Universe. Moreover, we will also assume that the gravitational field is sourced by the sum of energy densities of both components. We use the subscript t to refer to total quantities, ρ_t , P_t , δ_t etc., and the subscript X for a generic fluid, which may represent either matter or dark energy. So for perfect fluids we will have two equations for each fluid (for δ'_X and θ'_X), two equations for the gravitational field (i.e. for Φ and Φ'), and another one that provides the relation between Φ and Ψ .

The equations in the following are best obtained by an algebraic manipulator. In this case it is convenient to start directly with the perturbed flat-space metric in $N = \ln a$:

$$ds^2 = e^{2N} \left[-(1 + 2\Psi)\mathcal{H}^{-2}dN^2 + (1 + 2\Phi)\delta_{ij}dx^i dx^j \right], \quad (11.1)$$

and work out from the beginning a single mode k , putting $\Phi(r, a) = \Phi(a)e^{ik \cdot r}$, $\Psi(r, a) = \Psi(a)e^{ik \cdot r}$ etc. This new metric forces a new definition of the first-order four-velocity:

$$u^\alpha = \frac{dx^\alpha}{ds} = \left\{ \frac{dN}{a(1 + \Psi)\mathcal{H}^{-1}dN}, \frac{dx^i}{a\mathcal{H}^{-1}dN} \right\} \approx \left\{ \frac{\mathcal{H}}{a}(1 - \Psi), \frac{v^i}{a} \right\}, \quad (11.2)$$

where of course $a = e^N$ and $adN = da$. It is convenient also to define a new velocity divergence:

$$\theta_{\text{new}} = \frac{i\mathbf{k} \cdot \mathbf{v}}{\mathcal{H}} = \frac{\theta_{\text{old}}}{\mathcal{H}}, \quad (11.3)$$

where the old velocity divergence θ_{old} is defined in Eq. (4.63). For simplicity we drop the subscript “new” but will remind the reader of the new definition when necessary.

All the equations can be converted into conformal-time equations by using the rules

$$\frac{d}{dN} = \frac{1}{\mathcal{H}} \frac{d}{d\eta}, \quad (11.4)$$

$$\frac{d^2}{dN^2} = \frac{1}{\mathcal{H}^2} \frac{d^2}{d\eta^2} - \frac{d\mathcal{H}/d\eta}{\mathcal{H}^2} \frac{d}{d\eta}, \quad (11.5)$$

and into ordinary time equations by the same rules and replacing $\eta \rightarrow t$ and $\mathcal{H} \rightarrow H$.

If we have many fluids, the total energy-momentum tensor is the sum $T_{\alpha\beta} = \sum_i T_{(i)\alpha\beta}$ of the individual tensors. At the perturbation level this implies that [see Eqs. (4.37)–(4.39)]

$$\delta T_0^0 = -\rho_t \delta_t = -\sum_i \rho_i \delta_i, \quad (11.6)$$

$$ik_j(\delta T_0^j) = -(1 + w_{\text{eff}})\rho_t \theta_t = -\sum_i (1 + w_i)\rho_i \theta_i, \quad (11.7)$$

$$\delta T_1^1 = \delta T_2^2 = \delta T_3^3 = c_{s,t}^2 \rho_t \delta_t = \sum_i c_{s,i}^2 \rho_i \delta_i, \quad (11.8)$$

where the total perturbation variables are given by

$$\delta_t = \sum_i \Omega_i \delta_i, \quad (11.9)$$

$$\theta_t = \sum_i \frac{1 + w_i}{1 + w_{\text{eff}}} \Omega_i \theta_i, \quad (11.10)$$

together with the total equation of state and the sound speed

$$w_{\text{eff}} = \frac{P_t}{\rho_t} = \sum_i \Omega_i w_i, \quad (11.11)$$

$$c_{s,t}^2 = \frac{\sum_i c_{s,i}^2 \Omega_i \delta_i}{\delta_t} = \frac{\sum_i c_{s,i}^2 \Omega_i \delta_i}{\sum_i \Omega_i \delta_i}. \quad (11.12)$$

Recall that Ω_i is defined by $\Omega_i \equiv \rho_i / \rho_t$.

The total equation of state w_{eff} satisfies the following relation

$$\frac{\mathcal{H}'}{\mathcal{H}} = 1 + \frac{H'}{H} = -\frac{1}{2} - \frac{3}{2} w_{\text{eff}}. \quad (11.13)$$

The total sound speed simplifies if the i -th component is the only one to cluster ($\delta_i \neq 0$) since then $c_{s,t} = c_{s,i}$. If that component is also barotropic, i.e. $P_i = P_i(\rho_i)$, then the adiabatic sound speed is a function of w_i given by

$$c_{s(a),t}^2 = \frac{\dot{P}_i}{\dot{\rho}_i} = w_i - \frac{w'_i}{3(1 + w_i)}. \quad (11.14)$$

Suppose now all components are barotropic, $c_{s,i}^2 = dP_i/d\rho_i$. Under which condition is the total fluid barotropic? If we impose the adiabatic condition,

$$\frac{\delta \rho_i}{\rho'_i} = \frac{\delta \rho_j}{\rho'_j} \quad \rightarrow \quad \frac{\delta_i}{1 + w_i} = \frac{\delta_j}{1 + w_j}, \quad (11.15)$$

for different matter components i, j , one can express any δ_i as $\delta_1(1 + w_i)/(1 + w_1)$, where δ_1 corresponds to the perturbation for one component. Substituting this into Eq. (11.12), we find

$$\begin{aligned} c_{s,t}^2 &= \frac{\sum_i c_{s,i}^2 \Omega_i \delta_1 (1 + w_i)}{\sum_i \Omega_i \delta_1 (1 + w_i)} \\ &= \frac{\sum_i c_{s,i}^2 \rho_i (1 + w_i)}{\sum_i \rho_i (1 + w_i)} = \frac{\sum_i (dP_i/d\rho_i) \dot{\rho}_i}{\sum_i \dot{\rho}_i} = \frac{\dot{P}_t}{\dot{\rho}_t} = w_{\text{eff}} - \frac{w'_{\text{eff}}}{3(1 + w_{\text{eff}})}. \end{aligned} \quad (11.16)$$

Hence the total fluid remains barotropic provided that all components satisfy the adiabatic conditions. This occurs most notably on super-horizon scales for a Universe composed of dust and radiation, see Eq. (4.180). In general even if all the fluids are barotropic, the total fluid is not, or in other words $P_i = P_i(\rho_i)$ does not imply $P_t = P_t(\rho_t)$.

From Eqs. (4.61) and (4.62) the perturbation equations for a generic perfect fluid with density contrast δ_X and velocity divergence θ_X are given by

$$\delta'_X = 3(w - c_s^2)\delta_X - (\theta_X + 3\Phi')(1 + w), \quad (11.17)$$

$$\theta'_X = \left(3w - 1 - \frac{w'}{1 + w} - \frac{\mathcal{H}'}{\mathcal{H}}\right)\theta_X + \frac{c_s^2}{\hat{\lambda}^2(1 + w)}\delta_X + \frac{\Psi}{\hat{\lambda}^2}. \quad (11.18)$$

Here we have introduced the quantity

$$\hat{\lambda} \equiv \mathcal{H}/k = aH/k, \quad (11.19)$$

so as to check the dimensional correctness at once. Another advantage is that in real space we can interpret $\hat{\lambda}^{-2}$ as the operator $-\mathcal{H}^{-2}\nabla^2$ while in Fourier space, $\hat{\lambda} = \mathcal{H}/k$. In this way the perturbation equations can be read equivalently in real or Fourier space. Note that the above equations are valid for $w \neq -1$. From Eqs. (4.57) and (4.58) we obtain the following equations

$$\Phi = 3\hat{\lambda}^2 \left(\frac{1}{2}\delta_t + \Psi - \Phi' \right), \quad (11.20)$$

$$\Phi' = \Psi - \frac{3}{2}\hat{\lambda}^2\theta_t(1 + w_{\text{eff}}), \quad (11.21)$$

where we have used the background equation, $3\mathcal{H}^2 = 8\pi G a^2 \rho_t$. For $C_s^2 = w = -1$ the equations for δ_X and θ_X give rise to the solution $\delta_X = \theta_X = 0$, which means that the cosmological constant does not fluctuate. The perturbation equations for δ_X and θ_X are generic. For dark energy we would have $w = w(a)$, $c_s^2 = c_s^2(a)$; for cold dark matter $w = 0$, $c_s^2 = 0$; for radiation $w = c_s^2 = 1/3$, etc.

Equations (11.17) and (11.18) can be also applied to the *total* component, replacing the subscript X for t and w , c_s for w_{eff} , $c_{s,t}$. So the problem is composed of two equations for δ_X , θ_X , two for δ_t , θ_t , and two algebraic relations that couple them through Φ and Ψ . Any non-degenerate combination of four of these will be mathematically equivalent. These equations are therefore all we need for the general problem of several uncoupled perfect fluid components.

As we have already explained in Chapter 4, the (i, j) off-diagonal equations produce an additional equation for Φ and Ψ . In the absence of anisotropic stress this is simply given by $\Phi = -\Psi$. Using this identity, the gravitational equations

for Φ , Φ' can be written as

$$\Phi = \frac{3}{2}\hat{\lambda}^2[\delta_t + 3\hat{\lambda}^2\theta_t(1 + w_{\text{eff}})], \quad (11.22)$$

$$\Phi' = -\frac{3}{2}\hat{\lambda}^2[\delta_t + \theta_t(3\hat{\lambda}^2 + 1)(1 + w_{\text{eff}})]. \quad (11.23)$$

It is important to observe that a consequence of the gravitational gauge freedom is that only gauge-invariant quantities can be compared directly to observations. It is possible to show that a gauge-invariant combination reduces to [90]

$$\Delta_t \equiv \delta_t + 3\hat{\lambda}^2\theta_t(1 + w_{\text{eff}}), \quad (11.24)$$

in any gauge in which the $(0, i)$ elements [called w_i in Eq. (4.4)] of the perturbed metric are set to zero. These elements vanish for any observer at rest with respect to the coordinate frame and this is indeed what any observer assumes implicitly. The combination Δ_t is therefore the quantity to confront with observations. This reduces to the familiar δ_t only at small scales. It is therefore only in this limit that δ_t can be directly compared to the observed density contrast (at least in principle: in practice, there are a number of obstacles such as the problems of bias, of nonlinearities, of redshift distortions). From Eq. (11.22) we see that Δ_t essentially measures the total potential Φ . Generally speaking, we will discuss the evolution of δ_t only in the limit that $\hat{\lambda} \ll 1$. When this limit is not respected (e.g., when discussing CMB, ISW, lensing), we stick with Φ and Ψ .

The total variables δ_t and θ_t satisfy equations similar to (11.17) and (11.18) apart from a subscript t (δ_t , $c_{s,t}$ etc.) and w replaced by w_{eff} . Let us write the difference of Φ and Ψ in the form

$$\Phi = -\Psi + \sigma, \quad (11.25)$$

where $\sigma(k, t)$ is a generic function of space and time that represents the anisotropic stress. Differentiating Eq. (11.21) with respect to N and using Eq. (11.20) plus the equations for δ_t and θ_t , one obtains a relatively simple second-order equation for Φ :

$$\Phi'' + \left[3c_{s,t}^2 + \frac{1}{2}(5 - 3w_{\text{eff}}) \right] \Phi' + [(3 + \hat{\lambda}^{-2})c_{s,t}^2 - 3w_{\text{eff}}] \Phi = 3(c_{s,t}^2 - w_{\text{eff}})\sigma + \sigma'. \quad (11.26)$$

Note that we have already encountered this equation for $\sigma = 0$ and in conformal time in Eq. (4.66). If we switch off σ and also assume $c_{s,t}^2 = w_{\text{eff}}$ (which applies for a single fluid with constant w), then the third term on the l.h.s. of Eq. (11.26) vanishes in the large-scale limit ($\hat{\lambda} \gg 1$). This gives rise to a solution $\Phi = \text{constant}$, as we have already mentioned. This shows that the gravitational potential is constant

on super-horizon scales ($\hat{\lambda} \gg 1$) both for matter- and radiation-dominated regimes (but not during the transition!) and in general for any perfect fluid with constant w .

In the rest of this section we always assume $\sigma = 0$, unless otherwise stated. From Eq. (11.26) we see that Φ oscillates acoustically if $\hat{\lambda}^2 < c_{s,t}^2/[3(w_{\text{eff}} - c_{s,t}^2)]$. For larger scales, Φ grows if $c_{s,t}^2 < w_{\text{eff}}$ and decays otherwise. If the total sound speed is adiabatic, the equation for Φ at large scales becomes

$$\Phi'' + \left(\frac{5}{2} + \frac{3}{2}w_{\text{eff}} - \frac{w'_{\text{eff}}}{1 + w_{\text{eff}}} \right) \Phi' - \frac{w'_{\text{eff}}}{1 + w_{\text{eff}}} \Phi = 0, \quad (11.27)$$

where we have used Eq. (11.14). For a mixture of radiation and matter we have that $w_{\text{eff}} = \rho_r/[3(\rho_m + \rho_r)] = 1/[3(1 + a/a_{\text{eq}})]$, in which case one finds the exact solution (4.201).

For the Λ CDM model (which also implies $\sigma = 0$), the total sound speed after the radiation era is given by $c_{s,t}^2 = (\dot{P}_m + \dot{P}_{\text{DE}})/(\dot{\rho}_m + \dot{\rho}_{\text{DE}}) = \dot{P}_m/\dot{\rho}_m = 0$. As a consequence we immediately see that the evolution of Φ is completely *scale independent*. For the Λ CDM model we have $w_{\text{eff}} = -\Omega_{\text{DE}} = -1 + \Omega_m$ or

$$w_{\text{eff}} = \frac{\Omega_m^{(0)} - 1}{1 - \Omega_m^{(0)} + \Omega_m^{(0)} e^{-3N}}, \quad (11.28)$$

which goes as expected from 0 in the past to -1 in the future (note that the present epoch corresponds to $N = 0$). Inserting this into Eq. (11.26) one can directly solve the Φ equation numerically, see Fig. 4.1. The gravitational potential stays constant at early times, but it starts to decay after the Λ term dominates.

If one has to deal with a single component, the simplest way to proceed is to integrate Eq. (11.26) and then use Eqs. (11.17) and (11.18) to obtain δ_t and θ_t . For more general cases the Φ equation is not sufficient. For pressureless matter plus general dark energy we have

$$c_{s,t}^2 = \frac{\delta P_{\text{DE}} + \delta P_m}{\delta \rho_{\text{DE}} + \delta \rho_m} = \frac{c_{s,\text{DE}}^2 \Omega_{\text{DE}}}{\Omega_m(\delta_m/\delta_{\text{DE}}) + \Omega_{\text{DE}}} = c_{s,\text{DE}}^2 \Omega_{\text{DE}} \frac{\delta_{\text{DE}}}{\delta_t} = c_{s,\text{DE}}^2 \left(1 - \frac{\Omega_m \delta_m}{\delta_t} \right), \quad (11.29)$$

where $c_{s,\text{DE}}^2 = \delta P_{\text{DE}}/\delta \rho_{\text{DE}}$. Therefore we need to know the behavior of $\delta_{\text{DE}}/\delta_t$ or equivalently δ_m/δ_t . Another useful form, valid for $\Phi = -\Psi$, is

$$c_{s,t}^2 = c_{s,\text{DE}}^2 \left[1 - \frac{\Omega_m \delta_m}{2\Phi' + 2\Phi(1 + 1/(3\hat{\lambda}^2))} \right], \quad (11.30)$$

where we have used Eqs. (11.22) and (11.23). We proceed to derive a general second-order equation of δ for a generic perfect fluid component.

Differentiating Eq. (11.17) with respect to N and using Eq. (11.18), we obtain

$$\delta_X'' + a_1 \delta_X' + a_0 \delta_X = b_0(\Phi - \sigma) + b_1 \Phi' + b_2 \Phi'', \quad (11.31)$$

where

$$a_0 = \frac{1}{2} [c_s^2(3 + 2\hat{\lambda}^{-2} - 18w - 9w_{\text{eff}}) + 3w(-1 + 6w + 3w_{\text{eff}}) + 6(c_s^{2'} - w')] , \quad (11.32)$$

$$a_1 = \frac{1}{2}(1 + 6c_s^2 - 12w - 3w_{\text{eff}}) , \quad (11.33)$$

$$b_0 = \hat{\lambda}^{-2}(1 + w) , \quad (11.34)$$

$$b_1 = -\frac{3}{2} [2w' + (1 + w)(1 - 6w - 3w_{\text{eff}})] , \quad (11.35)$$

$$b_2 = -3(1 + w) . \quad (11.36)$$

This equation holds for each perfect fluid component in a multi-fluid medium. Since it has been obtained by manipulating only the conservation equations, and not the gravity sector, Eq. (11.31) applies also in any form of modified gravity that obeys the standard conservation laws. Let us remind the reader again that in this chapter the primes correspond to d/dN and that $\theta = i\mathbf{k} \cdot \mathbf{v}/\mathcal{H}$.

The combination of Eqs. (11.31) and (11.26) forms a closed set of equations if no anisotropic stress is present, to be supplemented only by the background solution and by the specified equations of state and sound speeds. For any additional perfect fluid component, we just need to add to the system another equation (11.31) with the specific w, c_s . Let us write it down for two cases, radiation and matter. Since the radiation corresponds to $w = c_s^2 = 1/3$, we obtain the following equation for the modes deep inside the Hubble radius ($\hat{\lambda} \ll 1$):

$$\delta_\gamma'' - \frac{1}{2}(1 + 3w_{\text{eff}})\delta_\gamma' + \frac{1}{3}\hat{\lambda}^{-2}\delta_\gamma = \frac{4}{3}\hat{\lambda}^{-2}\Phi + 2(1 + 3w_{\text{eff}})\Phi' - 4\Phi'' , \quad (11.37)$$

where we have set $\sigma = 0$. Since δ_γ is associated with the temperature anisotropy Θ_0 via the relation $\delta_\gamma = 4\Theta_0$, the conversion of Eq. (11.37) in terms of the derivative of the conformal time η leads to the following equation in the deep radiation era ($w_{\text{eff}} \simeq 1/3$):

$$\frac{d^2\Theta_0}{d\eta^2} + \frac{1}{3}k^2\Theta_0 = \frac{k^2}{3}\Phi - \frac{d^2\Phi}{d\eta^2} . \quad (11.38)$$

This corresponds to the $R_s \rightarrow 0$ limit of Eq. (5.13) with $c_s^2 = 1/3$ and $\Psi = -\Phi$.

For pressureless matter, Eq. (11.31) reduces to

$$\delta_m'' + \frac{1}{2}(1 - 3w_{\text{eff}})\delta_m' = -\hat{\lambda}^{-2}\Psi - \frac{3}{2}(1 - 3w_{\text{eff}})\Phi' - 3\Phi'' . \quad (11.39)$$

Equations (11.39) and (11.26) form a very general set of perturbation equations for any pressureless matter in the presence of dark energy. From these one can then derive algebraically all the other variables, δ_t , δ_{DE} , θ_m , etc.

At small scales $\hat{\lambda} \ll 1$ and for vanishing σ , Eqs. (11.39), (11.22), and (11.30) give

$$\delta_m'' + \frac{1}{2}(1 - 3w_{\text{eff}})\delta_m' = \frac{\Phi}{\hat{\lambda}^2}, \quad (11.40)$$

$$\Phi = \frac{3}{2}\hat{\lambda}^2\delta_t = \frac{3}{2}\hat{\lambda}^2(\Omega_m\delta_m + \Omega_{\text{DE}}\delta_{\text{DE}}), \quad (11.41)$$

$$c_{s,t}^2 = c_{s,\text{DE}}^2 \left[1 - \frac{3\hat{\lambda}^2\Omega_m\delta_m}{2\Phi} \right]. \quad (11.42)$$

If dark energy does not cluster then we have $\delta_{\text{DE}} = 0$, so that $\Phi = 3\hat{\lambda}^2\Omega_m\delta_m/2$ from Eq. (11.41). Equation (11.42) shows that in this case the total sound speed $c_{s,t}$ vanishes. From Eq. (11.40) it follows that

$$\delta_m'' + \frac{1}{2}(1 - 3\Omega_{\text{DE}}w_{\text{DE}})\delta_m' - \frac{3}{2}\Omega_m\delta_m = 0, \quad (11.43)$$

where we have used $w_{\text{eff}} = \Omega_{\text{DE}}w_{\text{DE}}$. This equation is completely fixed by assigning a $w_{\text{DE}}(a)$ given by the model and the present value $\Omega_m^{(0)} = 1 - \Omega_{\text{DE}}^{(0)}$ from which $\Omega_m(a) = \Omega_m^{(0)}a^{-3}/[\Omega_m^{(0)}a^{-3} + (1 - \Omega_m^{(0)})a^{-3(1+\hat{w}_{\text{DE}})}]$, with

$$\hat{w}_{\text{DE}}(N) = \frac{1}{N} \int_0^N w_{\text{DE}}(\tilde{N}) d\tilde{N}. \quad (11.44)$$

In some simple cases an analytical solution in terms of hypergeometric functions can be found [477]. It is often more useful however to work with an approximate solution. By using the growth rate parameter f defined in Eq. (4.107), Eq. (11.43) can be written in the form

$$f' + f^2 + \left[\frac{1}{2} - \frac{3}{2}w_{\text{DE}}(1 - \Omega_m) \right] f = \frac{3}{2}\Omega_m. \quad (11.45)$$

By using the Friedmann equation $3H^2 = 8\pi G(\rho_m + \rho_{\text{DE}})$ together with the continuity equation $\dot{\rho}_{\text{DE}} + 3H(1 + w_{\text{DE}})\rho_{\text{DE}} = 0$, we obtain

$$\Omega_m' = 3w_{\text{DE}}(1 - \Omega_m)\Omega_m. \quad (11.46)$$

Combining Eqs. (11.45) and (11.46), it follows that

$$3w_{\text{DE}}\Omega_m(1 - \Omega_m)\frac{df}{d\Omega_m} + \left[\frac{1}{2} - \frac{3}{2}w_{\text{DE}}(1 - \Omega_m) \right] f + f^2 = \frac{3}{2}\Omega_m. \quad (11.47)$$

Substituting $f = \Omega_m^\gamma$ into Eq. (11.47), we find that [478]

$$3w_{\text{DE}}\Omega_m(1-\Omega_m)(\ln \Omega_m) \frac{d\gamma}{d\Omega_m} - 3w_{\text{DE}} \left(\gamma - \frac{1}{2} \right) \Omega_m + \Omega_m^\gamma - \frac{3}{2}\Omega_m^{1-\gamma} + 3w_{\text{DE}}\gamma - \frac{3}{2}w_{\text{DE}} + \frac{1}{2} = 0. \quad (11.48)$$

If the variation of $w_{\text{DE}}(z)$ is slow so that the condition $|dw_{\text{DE}}/d\Omega_m| \ll 1/(1-\Omega_m)$ is satisfied, we obtain the following estimate for γ [479]:

$$\gamma = \frac{3(1-w_{\text{DE}})}{5-6w_{\text{DE}}} + \frac{3}{125} \frac{(1-w_{\text{DE}})(1-3w_{\text{DE}}/2)}{(1-6w_{\text{DE}}/5)^2(1-12w_{\text{DE}}/5)} (1-\Omega_m) + \mathcal{O}((1-\Omega_m)^2). \quad (11.49)$$

However since the present value of $1-\Omega_m$ is not really small, a better approximation for the second term in γ is [478]

$$\frac{3}{125} \frac{(1-w_{\text{DE}})(1-3w_{\text{DE}}/2)}{(1-6w_{\text{DE}}/5)^3} (1-\Omega_m). \quad (11.50)$$

Note that the Λ CDM model corresponds to $\gamma \simeq 6/11 \simeq 0.545$.

Another fit for γ is provided by [480]

$$\gamma = 0.545 + 0.05[1 + w_{\text{DE}}(z=1)]. \quad (11.51)$$

In Fig. 11.1 we show the behavior of the perturbation growth and the comparison with the fit (11.51).

It is useful to remark that while γ does not depend strongly on w_{DE} , the rate $f = \Omega_m^\gamma$ is significantly affected by w_{DE} . At any given z , in fact, the dark energy component is more important for higher w_{DE} (assuming a constant w_{DE} for simplicity). Larger Ω_{DE} means smaller Ω_m and hence $f(z)$ *decreases* with increasing w_{DE} , i.e. structures grow more slowly. If we assign the same initial amplitude to δ_m , say at $z \approx 1100$ as set by CMB observations, then we conclude that the present amplitude is smaller for larger w_{DE} . On the other hand we may instead know the present matter amplitude δ_m , for instance because we measure it through weak lensing or by estimating the bias. In this case we want our model to reproduce today's observations and consequently we normalize δ_m to today. This would imply that at any given z the linear fluctuation amplitude $\delta_m(z)$ was *higher* for larger w_{DE} . These considerations have an important impact on the estimates of the abundance of collapsed objects.

Finally, it is also important to consider the limits of the Ω_m^γ parametrization. Since Ω_m is usually contained between 0 and 1, f cannot pass from values larger than unity to values below. This rigidity in the parametrization could be a problem for the cases in which $f > 1$ in the past, cases that we encounter in coupled dark energy [481].

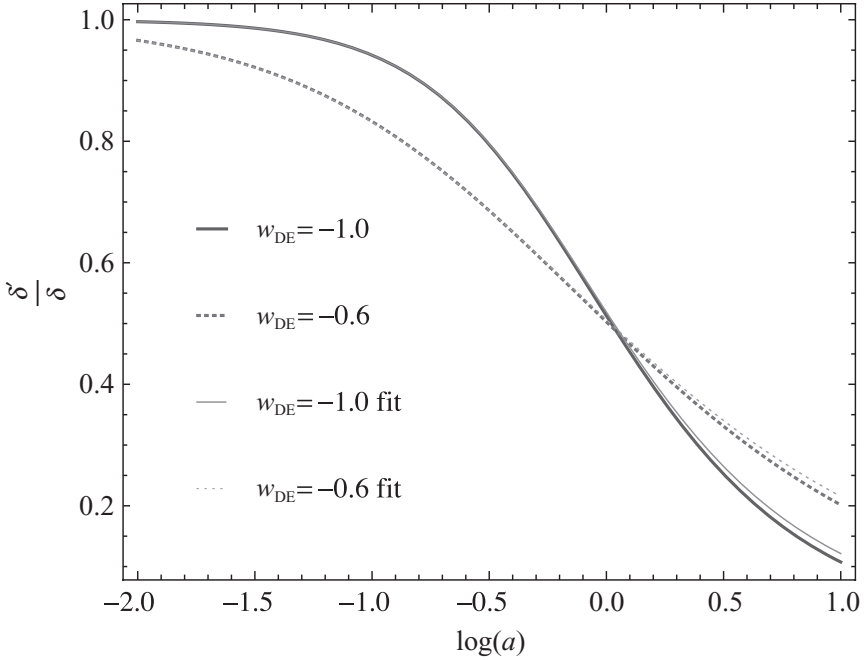


Figure 11.1 Growth rate $f = \delta'_m / \delta_m$ for $w_{\text{DE}} = -1$ (full line) and $w_{\text{DE}} = -0.6$ (dotted line) obtained by numerical integration compared to the approximation (11.51) (thin lines), fixing $\Omega_m^{(0)} = 0.3$.

Let us conclude this section by counting the degrees of freedom of our problem. For two perfect fluids (say, matter and dark energy) we have a complete system of two second-order equations (11.31), which makes four degrees of freedom. For any additional perfect fluid we should add two more degrees of freedom. It is possible to choose many other sets. For instance, in Ref. [482], the authors chose for two fluids (subscript 1, 2) gauge-invariant variables formed by Φ and three combinations of Φ , Φ' , δ_1 , δ_2 (think of fluid 1 as matter and of fluid 2 as dark energy). In our notation these variables are defined as

$$\mathcal{R} \equiv \Phi + \frac{2(\Phi + \Phi')}{3(1 + w_{\text{eff}})}, \quad (11.52)$$

$$S \equiv \Omega_2 \frac{(1 + w_2)\delta_1 - (1 + w_1)\delta_2}{1 + w_{\text{eff}}}, \quad (11.53)$$

$$\Gamma \equiv \frac{c_s^2 - c_{s(a)}^2}{1 - c_{s(a)}^2} \delta_2. \quad (11.54)$$

Here S is a gauge-invariant (relative) entropy perturbation, while Γ is an intrinsic entropy perturbation of the dark energy field 2. This set is particularly useful

to follow the behavior of S, Γ and to show that if one starts with $S = \Gamma = 0$ (adiabatic initial conditions) then S, Γ remain zero on large scales, *regardless* of the background evolution. For the system of non-relativistic matter and dark energy ($w_2 = w_{\text{DE}} \neq -1$), the condition $S = 0$ implies that $\delta_{\text{DE}} \approx \delta_m(1 + w_{\text{DE}})$ and hence

$$\begin{aligned} \delta_t &= \Omega_m \delta_m + \Omega_{\text{DE}} \delta_{\text{DE}} \\ &= \frac{\delta_m \Omega_m^{(0)}}{\Omega_m^{(0)} + (1 - \Omega_m^{(0)})a^{-3w_{\text{DE}}}} \left[1 + \frac{(1 - \Omega_m^{(0)})(1 + w_{\text{DE}})a^{-3w_{\text{DE}}}}{\Omega_m^{(0)}} \right]. \end{aligned} \quad (11.55)$$

Taking for instance $w_{\text{DE}} = -0.9$ we see that at $z \approx 1$ the contribution of dark energy fluctuations at super-horizon scales is near 3–4% of the matter one. This contribution can be searched for as an ISW effect on the CMB and in the cross-correlation with the large-scale structure, see Section 14.1.1.

11.2 Perturbations of a scalar field

As we have seen in Chapters 7 and 8, many models describe dark energy as a scalar field. In this section we derive perturbation equations for a quintessence scalar field. For generality, we include baryons and dark matter and assume explicit non-gravitational couplings between the field and the two matter components (i.e. coupled quintessence). Perturbations for scalar fields with non-canonical Lagrangians have been studied in e.g., Refs. [255, 483, 484, 485, 342].

Our dark energy model is therefore characterized by a general scalar-field potential $V(\phi)$ and general couplings $Q_i(\phi)$ to matter. Generalizing Eq. (8.71), the conservation equations with interacting terms for the field ϕ , cold dark matter (c), and baryons (b) are:

$$\nabla_\mu T_{(c)v}^\mu = Q_c(\phi) T_{(c)} \nabla_v \phi, \quad (11.56)$$

$$\nabla_\mu T_{(b)v}^\mu = Q_b(\phi) T_{(b)} \nabla_v \phi, \quad (11.57)$$

$$\nabla_\mu T_{(\phi)v}^\mu = -[Q_c(\phi) T_{(c)} + Q_b(\phi) T_{(b)}] \nabla_v \phi, \quad (11.58)$$

where the coupling functions $Q_{b,c}(\phi)$ depend on the species and T_i is the trace of the energy-momentum tensor of species i . Since radiation has a zero trace it is uncoupled to ϕ . As we have seen, this coupling form is motivated, through a conformal transformation, from Brans–Dicke gravity with species-dependent interaction. One could generalize the coupling in many ways, but this scalar-tensor form is sufficiently general for our purposes. The couplings are in general constrained by a number of observations. The baryon coupling in particular is severely constrained by local gravity experiments, unless the chameleon mechanism [321, 322] is at work (see the discussion in Chapter 9). Here for generality we leave the couplings completely free.

As we have already seen in Section 8.3.1, the field equations in the flat FLRW background are given by

$$\ddot{\phi} + 3H\dot{\phi} + V_{,\phi} = -(Q_c\rho_c + Q_b\rho_b), \quad (11.59)$$

$$\dot{\rho}_c + 3H\rho_c = Q_c\rho_c\dot{\phi}, \quad (11.60)$$

$$\dot{\rho}_b + 3H\rho_b = Q_b\rho_b\dot{\phi}, \quad (11.61)$$

$$\dot{\rho}_r + 4H\rho_r = 0, \quad (11.62)$$

$$3H^2 = \rho_\phi + \rho_c + \rho_b + \rho_r, \quad (11.63)$$

where we have used the unit $\kappa^2 = 1$. Sometimes the coupling constants $\beta_c = \sqrt{3/2}Q_c$ and $\beta_b = \sqrt{3/2}Q_b$ are introduced instead of Q_c and Q_b to simplify the background equations [17].

To study the perturbations we use the perturbed metric (11.1). In the following we neglect the contribution of radiation because it is unimportant when we discuss the perturbations after decoupling. If we define

$$u_\mu \equiv \frac{\phi_{,\mu}}{|g^{\alpha\beta}\phi_{,\alpha}\phi_{,\beta}|^{1/2}}, \quad \rho_\phi \equiv -\frac{1}{2}g^{\alpha\beta}\phi_{,\alpha}\phi_{,\beta} + V, \quad P_\phi \equiv -\frac{1}{2}g^{\alpha\beta}\phi_{,\alpha}\phi_{,\beta} - V, \quad (11.64)$$

the energy-momentum tensor of the scalar field can be written as

$$T_{(\phi)\mu\nu} = (\rho_\phi + P_\phi)u_\mu u_\nu + g_{\mu\nu}P_\phi. \quad (11.65)$$

In fact, the energy density and the pressure can be constructed as $\rho_\phi = T_{\mu\nu}u^\mu u^\nu$ and $P_\phi = T_{\mu\nu}h^{\mu\nu}$ with the help of the projection operator $h_{\mu\nu} = g_{\mu\nu} + u_\mu u_\nu$. We then obtain the following perturbations (notice here we use H not $\mathcal{H}$)

$$\delta\rho_\phi = H^2(\phi'\varphi' - \phi'^2\Phi) + V_{,\phi}\varphi, \quad (11.66)$$

$$\delta P_\phi = H^2(\phi'\varphi' - \phi'^2\Phi) - V_{,\phi}\varphi, \quad (11.67)$$

$$\theta_\phi \equiv -\frac{ik_i(\delta T_{0(\phi)}^i)}{(1 + w_\phi)\rho_\phi} = \hat{\lambda}^{-2}\frac{\varphi}{\phi'}, \quad (11.68)$$

where $V_{,\phi} \equiv dV/d\phi$ and

$$\varphi \equiv \delta\phi \quad (11.69)$$

denotes the field fluctuation. We also define the field density contrast

$$\delta_\phi \equiv \varphi/\phi. \quad (11.70)$$

The sound speed is therefore

$$c_{s,\phi}^2 = \frac{\delta P_\phi}{\delta\rho_\phi} = \frac{H^2(\phi'\varphi' - \phi'^2\Phi) - V_{,\phi}\varphi}{H^2(\phi'\varphi' - \phi'^2\Phi) + V_{,\phi}\varphi}. \quad (11.71)$$

If we put ourselves in the dark energy rest frame (that is, if we choose the gauge where dark energy is at rest), we have $\theta_\phi = 0$ and hence $\varphi = 0$. Then we see that the sound speed of the scalar-field rest frame equals unity.

We define also the dark energy effective mass

$$m_\phi^2 \equiv \frac{d^2 V}{d\phi^2}, \quad (11.72)$$

together with its dimensionless version

$$\hat{m}_\phi^2 \equiv m_\phi^2 / H^2. \quad (11.73)$$

Notice that in general m_ϕ is a function of ϕ . The mass becomes a constant only near the bottom of a harmonic potential. We also introduce the dimensionless potential

$$\hat{V} \equiv V / H^2, \quad (11.74)$$

which is at most of order unity. Perturbing the Einstein equations and the conservation equations, we obtain the linear perturbation equations below.

The perturbation equations for perfect fluids with generic equations of state $w_i = P_i / \rho_i$, couplings Q_i , and sound speeds $c_{s,i}$ are

$$\delta'_i = 3(w_i - c_{s,i}^2)(1 + 3Q_i\phi')\delta_i - (\theta_i + 3\Phi')(1 + w_i) + (1 - 3w_i)(Q_i\phi' + Q_{i,\phi}\phi'\phi), \quad (11.75)$$

$$\begin{aligned} \theta'_i = & -\frac{\theta_i}{2} \left[1 - 6w_i - 3w_{\text{eff}} + \frac{2w'_i}{1 + w_i} + 2Q_i(1 - 3w_i)\phi' \right] \\ & + \frac{1}{\hat{\lambda}^2} \left[\frac{c_{s,i}^2}{1 + w_i} \delta_i + \Psi + \frac{Q_i(1 - 3w_i)}{1 + w_i} \varphi \right], \end{aligned} \quad (11.76)$$

where $Q_{i,\phi} \equiv dQ_i/d\phi$. Note that these equations reduce to Eqs. (11.17) and (11.18) in the limit $Q_i \rightarrow 0$. For the models in which the equation of state is given by the form $w_i(\rho)$ instead of $w_i(a)$ (e.g., the Chaplygin gas model), the above equations are still valid with the substitution $c_{s,i}^2 \rightarrow w_i - w'_i/[3(1 + w_i)]$. The equation for the scalar field coupled to several fluids with equations of state $P_i = w_i(a)\rho_i$ and sound speeds $c_{s,i}$ is

$$\begin{aligned} \varphi'' + \left(2 + \frac{\mathcal{H}'}{\mathcal{H}} \right) \varphi' + (\hat{\lambda}^{-2} + \hat{m}_\phi^2) \varphi - \phi'(3\Phi' - \Psi') + 2\hat{V}_{,\phi} \Psi \\ = -3 \sum_i Q_i(1 - 3c_{s,i}^2) \Omega_i \delta_i - 6 \sum_i Q_i(1 - 3w_i) \Omega_i \Psi - 3 \sum_i (1 - 3w_i) Q_{i,\phi} \Omega_i \varphi. \end{aligned} \quad (11.77)$$

This equation for $Q_i = 0$ can be obtained also directly from the continuity equation (11.75) by making use of the perturbation variables in Eqs. (11.66)–(11.68) and putting $\delta_i = \delta_\phi$.

Finally the equations for metric perturbations are

$$\Phi = \frac{\hat{\lambda}^2}{2} \left[2\hat{V}\Psi + \varphi\hat{V}_{,\phi} + \phi'(3\varphi + \varphi') + 3 \sum \Omega_i \{ \delta_i + 3(1 + w_i)\hat{\lambda}^2\theta_i \} \right], \quad (11.78)$$

$$\Phi' = \frac{1}{2} \left[2\Psi - \varphi\phi' - 3\hat{\lambda}^2 \sum (1 + w_i)\theta_i \Omega_i \right], \quad (11.79)$$

plus the usual equivalence $\Phi = -\Psi$ in the absence of anisotropic stress.

Finally, in the uncoupled case with $\Phi = -\Psi$, the field perturbation equation (11.77) reduces to

$$\varphi'' + \left(2 + \frac{\mathcal{H}'}{\mathcal{H}} \right) \varphi' + (\hat{\lambda}^{-2} + \hat{m}_\phi^2)\varphi - 4\phi'\Phi' - 2\hat{V}_{,\phi}\Phi = 0. \quad (11.80)$$

Qualitatively, it is clear that one expects the scalar field to undergo damped oscillations for scales $\hat{\lambda} < 1/\hat{m}_\phi$. On these scales the scalar field will not contribute to the total gravitational potential and can be approximated as homogeneous. On larger scales the behavior depends on the mass term $\hat{m}_\phi$. For $\hat{m}_\phi < 1$ (i.e. $m_\phi < H$) both the background field ϕ and its perturbation φ evolve slowly. Then we can approximate $V_{,\phi} \simeq -3H^2\phi'$. Neglecting the ϕ' -dependent terms in Eq. (11.66), we obtain the following relation

$$\Omega_\phi\delta_\phi \simeq V_{,\phi}\varphi/(3H^2) \simeq -\phi'\varphi. \quad (11.81)$$

As a further approximation, we can take φ constant during slow-roll and approximate $\phi' \simeq [3(1 + w_\phi)\Omega_\phi]^{1/2}$ for w_ϕ near -1 . If w_ϕ remains constant then the field contribution $\Omega_\phi\delta_\phi$ to the total perturbation δ_i increases in proportion to $\Omega_\phi^{1/2}$ approximately.

For $\hat{m}_\phi > 1$ (i.e. $m_\phi > H$) the perturbations oscillate even on large scales. In this case, however, the background field will oscillate as well and the effective equation of state will depart from the one corresponding to dark energy. The field can now act indeed as dark matter and this case will be analyzed separately.

11.3 From dark energy to dark force

Let us now assume $\Phi = -\Psi$ and derive the perturbation solutions in the sub-horizon limit (small scales, $\hat{\lambda} \ll 1$). The gravitational equations are given by

$$\Phi = \frac{\hat{\lambda}^2}{2} \left[3 \sum \Omega_i \delta_i + \varphi\hat{V}_{,\phi} + \phi'(3\varphi + \varphi') \right], \quad (11.82)$$

$$\Phi' = -\frac{1}{2}\varphi\phi' - \Phi. \quad (11.83)$$

In the first equation we have used $\hat{\lambda}^2 \hat{V} \lesssim \hat{\lambda}^2 \ll 1$. Inserting Eqs. (11.82) and (11.83) into Eq. (11.77) and taking the small-scale limit, we obtain

$$\begin{aligned} \varphi'' + \left(2 + \frac{\mathcal{H}'}{\mathcal{H}}\right) \varphi' + \left[\hat{\lambda}^{-2} + \hat{m}_\phi^2 + 2\phi'^2 + 3 \sum (1 - 3w_i) \Omega_i Q_{i,\phi}\right] \varphi \\ \simeq -3 \sum Q_i (1 - 3c_{s,i}^2) \Omega_i \delta_i, \end{aligned} \quad (11.84)$$

where the sum is on the coupled components. Suppose now that (a) we can neglect the term $2\phi'^2\varphi$ since $|\phi'^2| (\lesssim 1)$ is much smaller than $\hat{\lambda}^{-2}$ and (b) we can assume also that the field potential is flat enough and its coupling is almost constant so that $\hat{m}_\phi^2$ and the term in $Q_{i,\phi}$ are negligible with respect to $\hat{\lambda}^{-2}$ (later we remove some of these approximations). In the limit of very small $\hat{\lambda}$ the field will undergo fast oscillations, forced by the term on the r.h.s. of Eq. (11.84). Averaging over the rapid oscillations of φ , we obtain

$$\langle \varphi \rangle \simeq -3\hat{\lambda}^2 \sum Q_i (1 - 3c_{s,i}^2) \Omega_i \delta_i. \quad (11.85)$$

Since the field is oscillating very fast, we must see this equation as giving the average of φ over many oscillations. This is the crucial difference between coupled and uncoupled fields concerning perturbations. In the coupled case the perturbed field φ does not oscillate around zero but acquires a non-zero average proportional to the couplings.

Since φ is of order $\hat{\lambda}^2$, Eq. (11.82) reduces to the usual Poisson equation

$$\Psi = -\frac{3}{2} \hat{\lambda}^2 \sum_i \Omega_i \delta_i. \quad (11.86)$$

Now, if we substitute $\langle \varphi \rangle$ into Eq. (11.76), we can define a new potential acting on the j -th component (which includes the effect of the coupling Q_j)

$$\begin{aligned} \hat{\Psi}_j &\equiv \Psi + \frac{Q_j(1 - 3w_j)}{1 + w_j} \langle \varphi \rangle \\ &= -\frac{3}{2} \hat{\lambda}^2 \sum_i \Omega_i \delta_i \left[1 + 2Q_i Q_j \frac{(1 - 3c_{s,i}^2)(1 - 3w_j)}{1 + w_j} \right]. \end{aligned} \quad (11.87)$$

Assuming for instance two matter components, CDM and baryons (subscripts c , b), we have a new potential on CDM:

$$\hat{\Psi}_c = -\frac{3}{2} \hat{\lambda}^2 \left[\Omega_b \delta_b (1 + 2Q_b Q_c) + \Omega_c \delta_c (1 + 2Q_c^2) \right]. \quad (11.88)$$

In real space, this equation becomes

$$\nabla^2 \hat{\Psi}_c = 4\pi G_{bc} \rho_b \delta_b + 4\pi G_{cc} \rho_c \delta_c, \quad (11.89)$$

where we have defined

$$G_{ij} = G\gamma_{ij}, \quad \gamma_{ij} \equiv 1 + 2Q_i Q_j. \quad (11.90)$$

Analogous equations hold for the baryon force equation.

We can now write down the sub-horizon linear equations for CDM and baryons. Since both Φ and φ are of the order of $\hat{\lambda}^2$, putting $w = c_s^2 = 0$ in Eqs. (11.75) and (11.76), we have

$$\delta'_c = -\theta_c, \quad (11.91)$$

$$\theta'_c = -\frac{1}{2}(1 - 3w_{\text{eff}} + 2Q_c\phi')\theta_c + \hat{\lambda}^{-2}\hat{\Psi}_c, \quad (11.92)$$

$$\delta'_b = -\theta_b, \quad (11.93)$$

$$\theta'_b = -\frac{1}{2}(1 - 3w_{\text{eff}} + 2Q_b\phi')\theta_b + \hat{\lambda}^{-2}\hat{\Psi}_b, \quad (11.94)$$

$$\hat{\Psi}_c = -\frac{3}{2}\hat{\lambda}^2(\gamma_{bc}\Omega_b\delta_b + \gamma_{cc}\Omega_c\delta_c), \quad (11.95)$$

$$\hat{\Psi}_b = -\frac{3}{2}\hat{\lambda}^2(\gamma_{bb}\Omega_b\delta_b + \gamma_{bc}\Omega_c\delta_c). \quad (11.96)$$

Differentiating Eq. (11.91) with respect to N and using Eq. (11.92), we obtain

$$\delta''_c + \frac{1}{2}(1 - 3w_{\text{eff}} + 2Q_c\phi')\delta'_c - \frac{3}{2}(\gamma_{cc}\delta_c\Omega_c + \gamma_{bc}\delta_b\Omega_b) = 0. \quad (11.97)$$

Similarly the equation for δ_b is given by

$$\delta''_b + \frac{1}{2}(1 - 3w_{\text{eff}} + 2Q_b\phi')\delta'_b - \frac{3}{2}(\gamma_{bb}\delta_b\Omega_b + \gamma_{bc}\delta_c\Omega_c) = 0. \quad (11.98)$$

These equations generalize the previous uncoupled equations (4.101) and (4.102).

Since baryons and dark matter obey different equations, they will develop a bias already at the linear level. A simple result can be obtained in the case where one component dominates. Assuming $\Omega_b \ll \Omega_c$, in fact, the baryon solution will be forced by the dominating CDM component to follow asymptotically its evolution. Defining the growth rate of δ_c as $f \equiv \delta'_c/\delta_c$ and putting $\delta_b = b\delta_c$ with $b = \text{constant}$, we obtain the coupled equations

$$f' + f^2 + \frac{1}{2}(1 - 3w_{\text{eff}} + 2Q_c\phi')f - \frac{3}{2}\gamma_{cc}\Omega_c = 0, \quad (11.99)$$

$$f' + f^2 + \frac{1}{2}(1 - 3w_{\text{eff}} + 2Q_b\phi')f - \frac{3}{2b}\gamma_{bc}\Omega_c = 0, \quad (11.100)$$

from which by subtraction

$$b = \frac{3\gamma_{bc}\Omega_c}{3\gamma_{cc}\Omega_c - 2(Q_c - Q_b)\phi'f}. \quad (11.101)$$

Notice that all terms on the r.h.s. are in general functions of time. This shows that a linear bias of gravitational nature develops whenever $Q_c \neq Q_b$. This bias extends to all sub-horizon scales and therefore is distinguishable from the hydrodynamical or non-linear bias that takes place in collapsed objects.

The growth rate f can be found numerically for any model by integrating (11.99). A simple analytical solution exists if w_{eff} , Q_c , Ω_c , ϕ' are constants (we are neglecting the baryons here):

$$f = -\frac{1}{4}(1 - 3w_{\text{eff}} + 2Q_c\phi') \pm \frac{1}{4}\sqrt{(1 - 3w_{\text{eff}} + 2Q_c\phi')^2 + 24\gamma_{cc}\Omega_c}. \quad (11.102)$$

This particular case occurs indeed on stationary solutions, e.g. the solution (d) in Section 8.3.1. In a pure matter-dominated cosmology we recover the standard solution $f = 1, -3/2$ for $w_{\text{eff}} = Q_c = 0$ and $\Omega_c = \gamma_{cc} = 1$. It is interesting to derive the limit of strong coupling for scaling solutions. This is obtained by the condition $Q \gg \lambda$ (λ is the potential slope) for the point (d) of Table 8.1. Then we have

$$\phi' = \sqrt{6}x_1 = 3/(Q_c + \lambda). \quad (11.103)$$

Inserting the values of ϕ' , w_{eff} , Ω_c for the point (d) into Eq. (11.102), we find that the growing mode solution corresponds to

$$f \simeq \sqrt{3Q_c\lambda}, \quad \text{for } Q_c \gg \lambda. \quad (11.104)$$

This diverges for $Q_c \rightarrow \infty$, which is due to the fact that in the limit of strong coupling the correction $1 + 2Q_c^2$ to gravity blows up. This shows that one can have fast-growing solutions in an accelerating Universe, even in the limit that $w_{\text{eff}} \rightarrow -1$. On the other hand this puts strong limits to the viability of scaling solutions since a fast growth during acceleration produces an excessive integrated Sachs–Wolfe effect [486].

Another simple case is the ϕ MDE scaling, i.e. the solution (a) in Section 8.3.1. Applying Eq. (11.102) to this case we obtain $f = 1 + 2Q_c^2$, which is faster than the standard CDM growth by $2Q_c^2$.

Finally, on accelerated but not scaling solutions and for small values of Q_c , it is also possible to find approximate solutions in the traditional form $f = \Omega_m^\gamma$, where [487]

$$\gamma \approx 0.55(1 - 2.6Q_c^2). \quad (11.105)$$

11.4 A massive dark energy field

It is interesting to observe that in Eq. (11.84) the terms in $Q_{i,\phi}$ contribute to the equation as effective masses. For a single pressureless fluid component (with a scalar field) this gives

$$\hat{m}_Q^2 \equiv 3\Omega Q_{,\phi}, \quad (11.106)$$

where Ω and Q are the density parameter and the coupling of the fluid, respectively. Let us consider the two effective masses of the dark energy field, previously neglected. If $\hat{\lambda}^{-2}$ is not much larger than $\hat{m}^2 = \hat{m}_\phi^2 + \hat{m}_Q^2$, Eq. (11.85) in Fourier space becomes (in this section we assume that dark energy is coupled to a single matter component, subscript m , or, equivalently, that has a universal coupling to all fields):

$$\varphi = -3Y(k)\hat{\lambda}^2 Q\Omega_m\delta_m, \quad (11.107)$$

where

$$Y(k) \equiv \frac{k^2}{k^2 + a^2 m^2}, \quad (11.108)$$

and $m = \hat{m}H$. If we substitute Eq. (11.107) into Eq. (11.76), we find that the effective potential is given by (neglecting the baryons)

$$\hat{\Psi} = -\frac{3}{2}\hat{\lambda}^2\Omega_m\delta_m[1 + 2Q^2Y(k)]. \quad (11.109)$$

Now, let us write down the present density in real space for a particle of mass M_0 located at the origin as $\rho_m^{(0)} = M_0\delta_D(0)$. The density contrast in “empty” space, i.e. for $\rho_m \gg \rho_t$, is therefore:

$$\Omega_m\delta_m = \frac{\rho_m}{\rho_t} = \frac{\kappa^2 a^3 \rho_m}{3\mathcal{H}^2 a} = \frac{\kappa^2 M(\phi)}{3\mathcal{H}^2 a}, \quad (11.110)$$

where

$$M(\phi) = \rho_m a^3 = M_0 e^{\int Q d\phi}. \quad (11.111)$$

Note that we have used the solution (8.92), i.e. $\rho_m = \rho_m^{(0)} \exp(\int Q d\phi) (a_0/a)^3$, together with $M_0 = \rho_m^{(0)} a_0^3 \delta_D(0)$. In Fourier space and for a unitary volume, the expression for the density contrast remains the same but the Dirac delta drops out. It turns out then that the potential originated by a dark matter particle in the linear regime is given by (we put back a k subscript for clarity)

$$\hat{\Psi}_k = -\frac{3}{2}\Omega_m\delta_m\hat{\lambda}^2[1 + 2Q^2Y(k)] = -4\pi G M(\phi) \left(\frac{1}{k^2} + \frac{2Q^2}{k^2 + a^2 m^2} \right) \frac{1}{a}, \quad (11.112)$$

from which we can define an effective G_{eff} in Fourier space:

$$G_{\text{eff}} = G \left(1 + \frac{2Q^2 k^2}{k^2 + a^2 m^2} \right). \quad (11.113)$$

Under the Fourier transformation

$$\hat{\Psi}(x) = \frac{1}{(2\pi)^3} \int e^{ik \cdot x} \hat{\Psi}_k d^3 k, \quad (11.114)$$

we have, using the angular integral

$$\int e^{ikx \cos \theta} \sin \theta d\theta d\phi = 4\pi \frac{\sin kx}{kx}, \quad (11.115)$$

the result

$$\begin{aligned} \hat{\Psi}(x) &= -\frac{2GM(\phi)}{\pi a} \int_0^\infty \frac{\sin kx}{kx} \left(1 + \frac{2Q^2 k^2}{k^2 + a^2 m^2} \right) dk \\ &= -\frac{2GM(\phi)}{\pi r} \int_0^\infty \frac{\sin y}{y} \left(1 + \frac{2Q^2 y^2}{y^2 + m^2 r^2} \right) dy, \end{aligned} \quad (11.116)$$

where $\mathbf{r} = a\mathbf{x}$ is the physical coordinate. The last integral gives finally the Yukawa potential

$$\hat{\Psi}(r) = -\frac{GM(\phi)}{r} (1 + 2Q^2 e^{-mr}). \quad (11.117)$$

Notice that in general both Q and m can be functions of ϕ and therefore of space and time. We have seen in Section 9.1.3 that the same Yukawa correction applies to $f(R)$ gravity ($Q = -1/\sqrt{6}$).

11.5 Sound speed of a scalar field

We know that an ultra-light scalar field behaves as a cosmological constant, at least as concerns the background dynamics. But what is the behavior at the perturbation level? And how can we understand the perturbation dynamics of a not-so-ultra-light scalar field?

The key quantity for this is the sound speed $c_{s,\phi}$. This sets the scale of clustering: if $c_{s,\phi}/\mathcal{H}$ is comparable to the horizon size then the field will not cluster on these scales; if this is small then the field might cluster. That is, we can have astrophysically sized ϕ fluctuations only if $c_{s,\phi} < 1$ (remember that velocities are in units of c). We have already seen in Section 11.2 that the sound speed for a scalar field is given by Eq. (11.71). Except for the case in which the potential is very flat, $V_{,\phi} \rightarrow 0$, for which $c_{s,\phi} = 1$, the field sound speed depends on the detailed behavior of both perturbation and background quantities. There is a limit

however which can be treated with some generality. Let us consider Eq. (11.84) for the *uncoupled* case

$$\varphi'' + \left(2 + \frac{\mathcal{H}'}{\mathcal{H}}\right) \varphi' + (\hat{\lambda}^{-2} + \hat{m}_\phi^2 + 2\phi'^2)\varphi = 0. \quad (11.118)$$

Notice that there is another “sound speed” here, namely the coefficient of the $\hat{\lambda}^{-2}\varphi$ term, always equal to unity. In Minkowski space this would be the velocity in the solution of the spacetime wave equation, $\varphi \sim \exp[ik(ct \pm r)]$. When this term dominates, the perturbed field φ oscillates acoustically and does not grow. However, we are not interested here in φ but rather in the density contrast $\delta\rho_\phi/\rho_\phi$.

Let us assume for a moment that $\hat{\lambda}$ and $\hat{m}_\phi$ are constant with a negligible contribution of ϕ'^2 . If the expansion rate H is also negligible, we can integrate the equation $\ddot{\varphi} + (k^2/a^2 + m_\phi^2)\varphi = 0$ to give the solution

$$\varphi = A \exp \left[\pm i m_\phi \left(1 + \frac{k^2}{2a^2 m_\phi^2} \right) t \right], \quad (11.119)$$

where we have assumed that $k^2 \ll a^2 m_\phi^2$.

Let us now approximate the potential V near its minimum as $V \approx \frac{1}{2} m_\phi^2 \phi^2$. If the field oscillates rapidly around the minimum we can assume that averaging over many oscillations the kinetic and the potential terms are equal:

$$\langle \dot{\phi}^2 \rangle \simeq m_\phi^2 \langle \phi^2 \rangle. \quad (11.120)$$

Notice that the kinetic term $\phi'^2 \simeq \langle \hat{V} \rangle \lesssim \Omega_{\text{DE}} \leq 1$ is then really negligible in Eq. (11.118) for sub-horizon modes ($\hat{\lambda} \lesssim 1$). Similarly, averaging over many oscillations, we also have

$$\langle \dot{\varphi}^2 \rangle \simeq m_\phi^2 \left(1 + \frac{k^2}{a^2 m_\phi^2} \right)^2 \langle \varphi^2 \rangle. \quad (11.121)$$

In fact, we can assume both ϕ and φ to be described by some sinusoidal $\sin(\mu t + \theta)$ where $\mu = m_\phi$ for ϕ and $\mu = m_\phi[1 + k^2/(2a^2 m_\phi^2)]$ for φ . Then averaging over many cycles we also have

$$\langle \dot{\phi}\dot{\varphi} \rangle \simeq m_\phi^2 \left(1 + \frac{k^2}{2a^2 m_\phi^2} \right) \langle \phi\varphi \rangle. \quad (11.122)$$

Since the Φ term in Eq. (11.71) is negligible for small scales, the sound speed squared is

$$c_{s,\phi}^2 = \frac{\delta P}{\delta \rho} \simeq \frac{\langle \dot{\phi}\dot{\varphi} \rangle - m_\phi^2 \langle \phi\varphi \rangle}{\langle \dot{\phi}\dot{\varphi} \rangle + m_\phi^2 \langle \phi\varphi \rangle} \simeq \frac{k^2}{4a^2 m_\phi^2}. \quad (11.123)$$

Therefore for $k \ll am_\phi$ we have

$$c_{s,\phi} = \frac{k}{2am_\phi} \ll 1. \quad (11.124)$$

The conclusion is that the k -th mode of the scalar field begins to cluster only for $k \ll am_\phi$. All this is in a linear regime, so this growth is observable only for scales which are still linear today or became non-linear only recently. The inverse mass $1/m_\phi$ sets the field Compton wavelength: only modes larger than $1/m_\phi$ can be localized and feel the pull of gravity. For masses near 10^{-28} eV the Compton wavelength m_ϕ^{-1} (that is, $\hbar/m_\phi c$) approximates the size of a galaxy. Such a field could therefore be considered as a form of dark matter rather than dark energy. In Ref. [488] such a field has been called “fuzzy dark matter.”

The mass scale m_ϕ sets also the timing for the onset of the oscillations. One could say that for a field slow-rolling over a quadratic potential the field begins oscillating when $m_\phi \gtrsim H$ and begins clustering on sub-horizon scales when $k \lesssim am_\phi$. If $m_\phi \lesssim H_0 \approx 10^{-33}$ eV neither of these two conditions is met yet and the field is not practically distinguishable from the cosmological constant. Then the Compton wavelength is as large as the present horizon scale of the Universe:

$$\frac{\hbar}{m_\phi} \gtrsim \frac{c}{H_0} = 2998 h^{-1} \text{ Mpc}. \quad (11.125)$$

In this case we can neglect altogether the fluctuation in the ϕ field on sub-horizon scales. However they are still relevant for super-horizon scales.

Finally, notice that there are cases in which the field is neither dark energy nor dark matter. The fields oscillating at the bottom of their potential with masses near $m_\phi \approx 10^{-30}$ eV cannot accelerate the present expansion (and therefore are not candidates for dark energy) nor can they clump into galaxies (and therefore are not candidates for dark matter). However they might still exist as a minor component of the Universe. The phenomenon is very similar to the massive neutrino case, in which an effective sound speed $c_s \simeq T_v^{(0)}/(am_v)$ can be defined when the neutrino becomes non-relativistic, where $T_v^{(0)}$ is the present neutrino temperature.

This intermediate massive component can cluster on sub-horizon scales larger than a Jeans (or “free-streaming”) scale corresponding to the wavenumber [see Eq. (4.79)]

$$k_J = \sqrt{\frac{3}{2}} \frac{\mathcal{H}}{c_{s,\phi}} = \sqrt{\frac{3}{2}} \frac{2a^2 m_\phi H}{k_J}. \quad (11.126)$$

During the matter-dominated epoch we have $H^2 = H_0^2 \Omega_m^{(0)} a^{-3}$, so that k_J is given by

$$k_J(a) \approx a^{1/4} m_\phi^{1/2} H_0^{1/2} \Omega_m^{(0)1/4}. \quad (11.127)$$

Each scale k therefore grows from the time when the field starts oscillating, given by the oscillation condition $H \simeq m_\phi$ or $a_{\min} \simeq (H_0^2 \Omega_m^{(0)} / m_\phi^2)^{1/3}$, until the value $a_{\max}$ such that $k = k_J(a_{\max})$, at which point the field enters the free-streaming regime. At this epoch, the field energy density stops sourcing the matter fluctuations and the growth is slowed down as in Eq. (4.105). Therefore we expect that between $a_{\min}$ and $a_{\max}$ matter perturbations on the scales $k > k_J$ grow slower than those at larger scales. This will give rise to a drop in power in the matter power spectrum. The larger the fraction Ω_ϕ is, the stronger the drop is, because more perturbations stop growing. This break in the spectrum can be quantified and compared to observations. The result is that a density fraction of a few percent of an intermediate massive field is allowed by observations [489].

11.6 Perturbations in modified gravity models

So far we have discussed the evolution of matter density perturbations in the framework of General Relativity. In this section we shall extend our analysis to modified gravity models of dark energy. This includes models such as $f(R)$ gravity, scalar-tensor gravity, and DGP braneworld models.

11.6.1 $f(R)$ gravity

First of all, let us consider the case of $f(R)$ gravity in the metric formalism in the presence of non-relativistic matter. The equations for matter perturbations are given by Eqs. (11.17) and (11.18) with $w_m = c_s^2 = 0$:

$$\delta'_m = -(\theta_m + 3\Phi') , \quad (11.128)$$

$$\theta'_m = \Psi / \hat{\lambda}^2 - (1 + \mathcal{H}' / \mathcal{H}) \theta_m . \quad (11.129)$$

Combining these equations gives Eq. (11.39) that we rewrite here in this form:

$$\delta''_m + \left(1 + \frac{\mathcal{H}'}{\mathcal{H}}\right) \delta'_m + \frac{1}{\hat{\lambda}^2} \Psi = -3 \left(1 + \frac{\mathcal{H}'}{\mathcal{H}}\right) \Phi' - 3\Phi'' . \quad (11.130)$$

For the modes deep inside the Hubble radius ($\hat{\lambda} \ll 1$) the r.h.s. of Eq. (11.130) can be neglected relative to the l.h.s., i.e.

$$\delta''_m + \left(1 + \frac{\mathcal{H}'}{\mathcal{H}}\right) \delta'_m + \frac{1}{\hat{\lambda}^2} \Psi = 0 . \quad (11.131)$$

In $f(R)$ gravity the quantity $F(R) = \partial f / \partial R$ has a perturbation δF . Perturbing Eq. (9.2), we obtain the following equations in Fourier space (in the unit of

$$\kappa^2 = 1) \text{ [417, 371]}$$

$$\begin{aligned} & -\frac{k^2}{a^2}\Phi + 3H(H\Psi - \dot{\Phi}) \\ & = \frac{1}{2F} \left[3H\delta\dot{F} - \left(3\dot{H} + 3H^2 - \frac{k^2}{a^2} \right) \delta F - 3H\dot{F}\Psi - 3\dot{F}(H\Psi - \dot{\Phi}) - \delta\rho_m \right], \end{aligned} \quad (11.132)$$

$$\begin{aligned} \delta\ddot{F} + 3H\delta\dot{F} + \left(\frac{k^2}{a^2} - \frac{R}{3} \right) \delta F &= \frac{1}{3}\delta\rho_m + \dot{F}(3H\Psi + \dot{\Psi} - 3\dot{\Phi}) \\ &+ (2\ddot{F} + 3H\dot{F})\Psi - \frac{1}{3}F\delta R, \end{aligned} \quad (11.133)$$

$$\Psi + \Phi = -\frac{\delta F}{F}, \quad (11.134)$$

where we used the time derivative with respect to cosmic time t . For the modes deep inside the Hubble radius, the terms including k^2/a^2 and $\delta\rho_m$ in Eq. (11.132) are the dominant contributions. Hence we obtain the following approximate relations from Eqs. (11.132) and (11.134):

$$\Phi = \frac{1}{2F} \left(\frac{a^2}{k^2} \delta\rho_m - \delta F \right), \quad \Psi = -\frac{1}{2F} \left(\frac{a^2}{k^2} \delta\rho_m + \delta F \right). \quad (11.135)$$

Provided that $|\dot{F}| \lesssim |HF|$ and $|\ddot{F}| \lesssim |H^2F|$, the second and third terms on the r.h.s. of Eq. (11.133) are much smaller than $\delta\rho_m$ and $(k^2/a^2)\delta F$ for the modes deep inside the Hubble radius. Using the relation $\delta R = \delta F/f_{,RR}$, we find that Eq. (11.133) is approximately given by

$$\delta\ddot{F} + 3H\delta\dot{F} + \left(\frac{k^2}{a^2} + M^2 \right) \delta F = \frac{1}{3}\delta\rho_m, \quad (11.136)$$

where

$$M^2 \equiv \frac{f_{,R}}{3f_{,RR}}. \quad (11.137)$$

In order to derive Eq. (11.136) we have used the following condition

$$\left\{ \frac{k^2}{a^2}, M^2 \right\} \gg R \sim H^2. \quad (11.138)$$

Note that the condition $M^2 \gg R$ is satisfied for viable $f(R)$ models in the past cosmic expansion history of the Universe [369, 371], see Eq. (9.52). From the stability of cosmological perturbations we require that $M^2 > 0$, which gives the condition $f_{,RR} > 0$ (provided that $f_{,R} > 0$). In the following we shall discuss two cases: (A) $M^2 \gg k^2/a^2$ and (B) $M^2 \ll k^2/a^2$, separately. In terms of the

characteristic function $m(r)$ introduced in Eq. (9.21) the conditions (A) and (B) can be written as $m \ll \hat{\lambda}^2$ and $m \gg \hat{\lambda}^2$, respectively. We recall that for viable $f(R)$ models the mass squared M^2 is large in the past and gradually decreases toward the present. Hence the transition from the region (A) to (B) can occur in the past, depending on the modes k .

(i) Evolution of perturbations in the regime $M^2 \gg k^2/a^2$.

The general solutions for Eq. (11.136) are given by the sum of the oscillating solution δF_{osc} obtained by setting $\delta\rho_m = 0$ and the special solution δF_{ind} of Eq. (11.136) induced by the presence of matter perturbations $\delta\rho_m$. The oscillating part δF_{osc} satisfies the equation $(a^{3/2}\delta F_{\text{osc}})'' + M^2(a^{3/2}\delta F_{\text{osc}}) \simeq 0$. By using the WKB approximation, we obtain the solution

$$\delta F_{\text{osc}} \propto a^{-3/2} f_{,RR}^{1/4} \cos\left(\int \frac{1}{\sqrt{3f_{,RR}}} dt\right). \quad (11.139)$$

Note that we have used $f_{,R} \simeq 1$ because the viable $f(R)$ models are close to the Λ CDM model when the mass M is heavy.

In the following let us consider the model (9.78) that corresponds to the asymptotic form of the models (9.9) and (9.10) in the region $R \gg R_c$. During the matter era in which the background Ricci scalar evolves as $R^{(0)} = 4/(3t^2)$, the quantity $f_{,RR}$ has a dependence $f_{,RR} \propto R^{-2(n+1)} \propto t^{4(n+1)}$. Hence the evolution of the perturbation, $\delta R_{\text{osc}} = \delta F_{\text{osc}}/f_{,RR}$, is given by

$$\delta R_{\text{osc}} \simeq c t^{-(3n+4)} \cos(c_0 t^{-2(n+1)}), \quad (11.140)$$

where c and c_0 are constants. As we go back to the past the perturbation δR_{osc} dominates over $R^{(0)} \propto t^{-2}$, unless the coefficient c is chosen to be very small. Since the Ricci scalar R can be negative in this case, this can lead to the violation of the stability conditions ($f_{,RR} > 0$ and $f_{,R} > 0$).

The special solution δF_{ind} of Eq. (11.136) can be derived by neglecting the first and second terms relative to others, giving

$$\delta F_{\text{ind}} \simeq \frac{\delta\rho_m}{3M^2}, \quad \delta R_{\text{ind}} \simeq \delta\rho_m. \quad (11.141)$$

Under the condition $|\delta F_{\text{osc}}| \ll |\delta F_{\text{ind}}|$ we have $\delta F \simeq \delta\rho_m/(3M^2)$, so that Eq. (11.135) reduces to

$$\Psi = -\Phi = -\frac{1}{2F} \frac{a^2}{k^2} \delta\rho_m. \quad (11.142)$$

Substituting Eq. (11.142) into Eq. (11.131), we find that the matter perturbation obeys the following equation

$$\delta_m'' + \left(1 + \frac{\mathcal{H}'}{\mathcal{H}}\right) \delta_m' - \frac{3}{2} \Omega_m \delta_m = 0, \quad (11.143)$$

where $\Omega_m = \rho_m/(3FH^2)$. This is the same form as the usual equation of matter perturbations in General Relativity, which has the growing solution

$$\delta_m \propto t^{2/3}. \quad (11.144)$$

From Eq. (11.141) we get

$$\delta F_{\text{ind}} \propto t^{4(n+2/3)}, \quad \delta R_{\text{ind}} \propto t^{-4/3}. \quad (11.145)$$

Compared to the oscillating mode (11.140), the induced matter mode δR_{ind} decreases more slowly and thus dominates in the late Universe. Relative to the background value $R^{(0)}$, the perturbation, $\delta R = \delta R_{\text{osc}} + \delta R_{\text{ind}}$, evolves as

$$\frac{\delta R}{R^{(0)}} \simeq b_1 t^{-(3n+2)} \cos(c_0 t^{-p}) + b_2 t^{2/3}, \quad (11.146)$$

where b_1 and b_2 are constants. In order to avoid the dominance of the oscillating mode at the early epoch, the coefficient b_1 needs to be suppressed relative to b_2 . Note that this property also persists for the evolution of matter perturbations during the radiation-dominated epoch [369, 371].

(ii) Evolution of perturbations in the regime $M^2 \ll k^2/a^2$.

Since the scalaron mass decreases as $M \propto t^{-2(n+1)}$, the modes that initially exist in the region $M^2 \gg k^2/a^2$ can enter the regime $M^2 \ll k^2/a^2$ during the matter-dominated epoch. It is sufficient to discuss the matter-induced mode because the oscillating mode is already suppressed during the evolution in the regime $M^2 \gg k^2/a^2$. The matter-induced special solution of Eq. (11.136) in the regime $M^2 \ll k^2/a^2$ is approximately given by

$$\delta F_{\text{ind}} \simeq \frac{a^2}{3k^2} \delta \rho_m. \quad (11.147)$$

From Eq. (11.135) the gravitational potentials satisfy

$$\Psi = -\frac{4}{3} \cdot \frac{1}{2F} \frac{a^2}{k^2} \delta \rho_m, \quad \Phi = \frac{2}{3} \cdot \frac{1}{2F} \frac{a^2}{k^2} \delta \rho_m. \quad (11.148)$$

Plugging Eq. (11.148) into Eq. (11.131), it follows that

$$\delta_m'' + \left(1 + \frac{\mathcal{H}'}{\mathcal{H}}\right) \delta_m' - \frac{4}{3} \cdot \frac{3}{2} \Omega_m \delta_m \simeq 0. \quad (11.149)$$

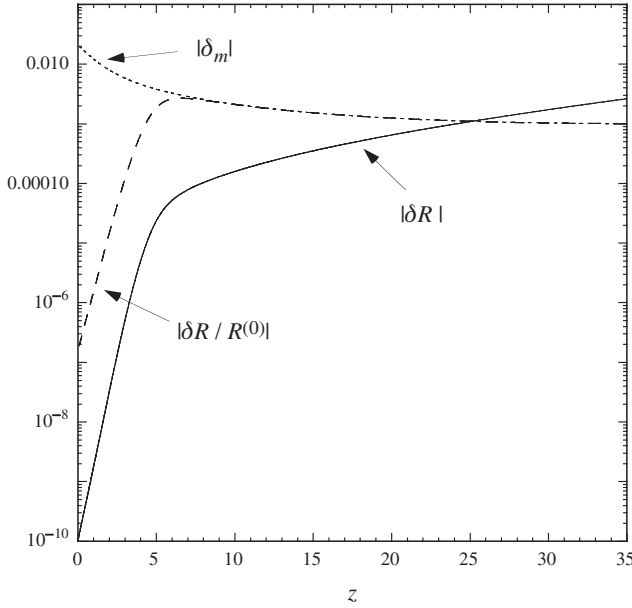


Figure 11.2 The evolution of δR , $\delta R/R^{(0)}$, and δ_m for the model (9.78) with $n = 1$ for the mode $k/(a_0 H_0) = 335$. This corresponds to the case in which the coefficient b_1 in Eq. (11.146) is small so that the oscillating mode δR_{osc} is negligible relative to the matter-induced mode δR_{ind} . In this case the transition from the region $M^2 \gg k^2/a^2$ to the region $M^2 \ll k^2/a^2$ occurs around the redshift $z = 5$. From Ref. [371].

Notice that the factor $4/3$ in the last term can be simply understood as the value of the Yukawa correction (11.113) in the limit of large k and for $Q = -1/\sqrt{6}$.

During the matter era with $\Omega_m \simeq 1$ and $a \propto t^{2/3}$, the matter perturbation evolves as

$$\delta_m \propto t^{\frac{\sqrt{33}-1}{6}}. \quad (11.150)$$

The growth rate of δ_m is larger than that in the region $M^2 \gg k^2/a^2$.

Let us consider the model (9.78). During the matter era we obtain $\delta F_{\text{ind}} \propto t^{\frac{\sqrt{33}-5}{6}}$ from Eqs. (11.147) and (11.150), and hence

$$\delta R_{\text{ind}} \propto t^{-4n + \frac{\sqrt{33}-29}{6}}, \quad \frac{\delta R_{\text{ind}}}{R^{(0)}} \propto t^{-4n + \frac{\sqrt{33}-17}{6}}. \quad (11.151)$$

In Fig. 11.2 we plot the evolution of δR , $\delta R/R^{(0)}$ and δ_m for the model (9.78) with $n = 1$ for the mode $k/(a_0 H_0) = 335$ (corresponding to roughly $2\pi/k \approx 60h^{-1}$ Mpc, i.e. a scale that is well within a linear regime). Note that we have chosen initial conditions so that the oscillating mode is negligible relative to the matter-induced mode. Initially the perturbation is in the region $M^2 \gg k^2/a^2$ and hence

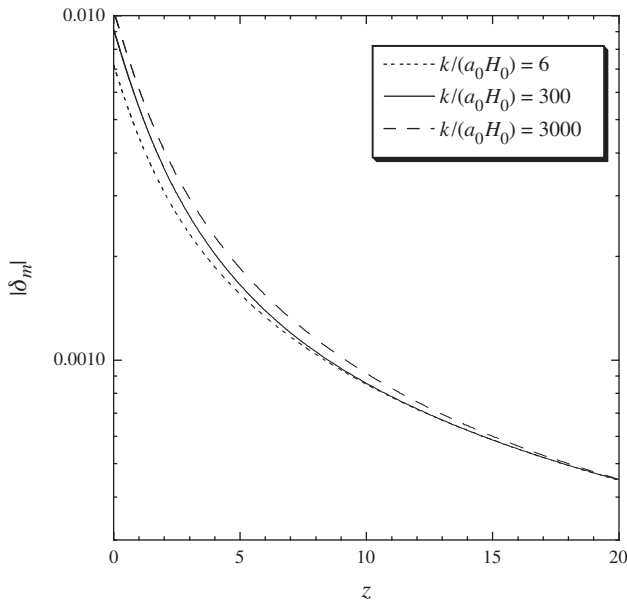


Figure 11.3 The evolution of the matter perturbation δ_m for the model (9.78) with $n = 0.5$. The curves correspond to $k/(a_0 H_0) = 6, 300, 3000$. The transition redshift z_k increases for larger k . For the mode $k = 300a_0 H_0$, $z_k = 9.6$. Non-linear effects are not taken into account. From Ref. [371].

the evolution of δ_m and $\delta R/R^{(0)}$ is given by Eqs. (11.144) and (11.146) respectively. The sudden decrease of $\delta R/R^{(0)}$ means that the system enters the region $M^2 \ll k^2/a^2$ in which the evolution of perturbations is given by Eqs. (11.150) and (11.151).

(iii) Matter power spectra

The evolution of the matter perturbation is given by $\delta_m \propto t^{2/3}$ for $M^2 \gg k^2/a^2$ and $\delta_m \propto t^{(\sqrt{33}-1)/6}$ for $M^2 \ll k^2/a^2$. We shall use the subscript “ k ” for the quantities at which k is equal to aM , whereas the subscript “ Λ ” is used for the onset of cosmic acceleration ($\ddot{a} = 0$). While the redshift z_Λ is independent of k , z_k depends on k and also on the mass M .

For the model (9.78) the variable $m = Rf_{,RR}/f_{,R}$ can grow fast from the regime $m \ll \hat{\lambda}^2$ (i.e. $M^2 \gg k^2/a^2$) to the regime $m \gg \hat{\lambda}^2$ (i.e. $M^2 \ll k^2/a^2$). In fact we recall that m can grow to as large as the order of 0.1 even if m is much smaller than 10^{-9} in the deep matter era. If the transition characterized by the condition $M^2 = k^2/a^2$ occurs during the deep matter era ($z \gg 1$), one can estimate the critical redshift z_k . We use the asymptotic forms $m \simeq C(-r-1)^{2n+1}$ with $C = 2n(2n+1)/\mu^{2n}$ and $r \simeq -1 - \mu R_c/R$ as well as the approximate relations $H^2 \simeq H_0^2 \Omega_m^0 (1+z)^3$ and $R \simeq 3H^2$. The present value of the dark energy density may be

approximated as $\rho_{\text{DE}}^{(0)} \approx \mu R_c/2$. Hence we have that $\mu R_c \approx 6H_0^2 \Omega_{\text{DE}}^{(0)}$, where $\Omega_{\text{DE}}^{(0)}$ is the density parameter of dark energy today. Then the condition $M^2 = k^2/a^2$, i.e. $m \simeq (aH/k)^2$, translates into the critical redshift

$$z_k \simeq \left[\left(\frac{k}{a_0 H_0} \right)^2 \frac{2n(2n+1)}{\mu^{2n}} \frac{(2\Omega_{\text{DE}}^{(0)})^{2n+1}}{\Omega_m^{(0)2(n+1)}} \right]^{\frac{1}{6n+4}} - 1. \quad (11.152)$$

For $n = 1$, $\mu = 3$, $k = 300a_0 H_0$, and $\Omega_m^{(0)} = 0.28$ the numerical value for the critical redshift is $z_k = 4.5$, which shows good agreement with the analytical value estimated by Eq. (11.152). We caution, however, that Eq. (11.152) begins to lose its accuracy for z_k close to 1. Equation (11.152) shows that z_k tends to be smaller for larger n and μ .

As n gets larger, the period of a non-standard evolution of δ_m becomes shorter. Since the scalaron mass evolves as $M \propto t^{-2(n+1)}$ for the model (9.78), the time t_k has a scale-dependence $t_k \propto k^{-\frac{3}{6n+4}}$. This means that the smaller-scale modes cross the transition point earlier. The matter power spectrum $P_{\delta_m} = |\delta_m|^2$ at the time t_Λ shows a difference compared to the case of the Λ CDM model:

$$\frac{P_{\delta_m}(t_\Lambda)}{P_{\delta_m}^{\Lambda\text{CDM}}(t_\Lambda)} = \left(\frac{t_\Lambda}{t_k} \right)^{2\left(\frac{\sqrt{33}-1}{6} - \frac{2}{3}\right)} \propto k^{\frac{\sqrt{33}-5}{6n+4}}. \quad (11.153)$$

While the galaxy matter power spectrum is modified by this effect, the CMB spectrum is hardly affected except for low multipoles around which the ISW effect becomes important. Thus there is a difference for the spectral indices of two power spectra, i.e.

$$\Delta n(t_\Lambda) = \frac{\sqrt{33}-5}{6n+4}. \quad (11.154)$$

For larger n the redshift z_k can be as close as z_Λ , which means that the estimation (11.154) is not necessarily valid in such cases. One finds that the estimation (11.154) agrees well with the numerically obtained $\Delta n(t_\Lambda)$ for $n \lesssim 2$ [371].

We note that the estimation (11.154) does not take into account the evolution of δ_m after $z = z_\Lambda$ to the present epoch ($z = 0$). After the system enters the epoch of cosmic acceleration, the momentum k can again become smaller than aM . Hence the k -dependence is not necessarily negligible even for $z < z_\Lambda$. However, numerical simulations show that $\Delta n(t_0)$ is not much different from $\Delta n(t_\Lambda)$ derived by Eq. (11.154) [371]. Thus the analytic estimation (11.154) is certainly reliable to place constraints on model parameters except for $n \gg 1$. Observationally we do not find any strong difference for the slopes of the spectra of LSS and CMB. If we take the mild bound $\Delta n(t_\Lambda) \lesssim 0.05$, we obtain the constraint $n \gtrsim 2$. In this case local gravity constraints are also satisfied, see Eq. (9.89).

The modified growth of matter perturbations also affects the evolution of the gravitational potentials Ψ and Φ . As we have seen the effective potential $\psi = \Phi - \Psi$ is important in discussing the ISW effect on the CMB as well as the weak lensing observations, see Eqs. (4.236) and (4.238). From Eq. (11.135) this potential is given by

$$\psi = \frac{3a^2 H^2}{k^2} \Omega_m \delta_m. \quad (11.155)$$

In the Λ CDM model the potential ψ remains constant during the standard matter era, but it decays after the system enters the accelerated epoch, producing the ISW contribution for low multipoles on the CMB power spectrum. In $f(R)$ gravity the additional growth of matter perturbations in the region $z < z_k$ changes the evolution of ψ .

From CMB observations, however, we do not obtain a constraint on n tighter than the one derived by the spectral index of matter perturbations [490]. This comes from the fact that the ISW effect is important only for the modes with $k/(a_0 H_0) = \mathcal{O}(1)$ whose transition redshift z_k is smaller than the modes relevant to the galaxy power spectrum. In the weak lensing observations, the modified evolution of the lensing potential ψ directly leads to the change even for the small-scale shear power spectrum [491, 492]. Hence this can be a powerful tool to constrain $f(R)$ gravity models from future observations.

11.6.2 Scalar-tensor gravity

Let us next discuss the case of scalar-tensor gravity. To be concrete we shall study the evolution of matter perturbations for the Jordan frame action (9.120), i.e. Brans–Dicke theory with the potential $U(\phi)$ and the coupling $F(\phi) = e^{-2Q\phi}$. We define the field mass squared to be

$$M^2 \equiv \frac{d^2 U}{d\phi^2}. \quad (11.156)$$

If the scalar field is light such that the condition $M \lesssim H_0$ is always satisfied irrespective of high- or low-density regions, the coupling Q is constrained to be $|Q| \lesssim 10^{-3}$ from local gravity tests. Meanwhile, if the mass M in the region of high density is much larger than that on cosmological scales, it is possible to satisfy local gravity constraints by the chameleon mechanism even if $|Q|$ is of the order of unity. Cosmologically the mass M can decrease from the past to the present, which can allow the transition from the “GR regime” to the “scalar-tensor regime” as happens in $f(R)$ gravity. An example of the field potential showing this behavior is given by Eq. (9.148).

As in $f(R)$ gravity, the matter perturbation δ_m satisfies Eq. (11.131). The difference appears in the expression of the gravitational potential Ψ . In Fourier space the scalar metric perturbations obey the following equations [417, 273]

$$\begin{aligned}
 -\frac{k^2}{a^2}\Phi + 3H(H\Psi - \dot{\Phi}) = & -\frac{1}{2F}\left[\omega\dot{\phi}\dot{\phi} + \frac{1}{2}(\omega_{,\phi}\dot{\phi}^2 - F_{,\phi}R + 2U_{,\phi})\varphi \right. \\
 & + \left(3\dot{H} + 3H^2 - \frac{k^2}{a^2}\right)\delta F - 3H\delta\dot{F} \\
 & \left. + (3H\dot{F} - \omega\dot{\phi}^2)\Psi + 3\dot{F}(H\Psi - \dot{\Phi}) + \delta\rho_m\right], \quad (11.157)
 \end{aligned}$$

$$\begin{aligned}
 \ddot{\varphi} + \left(3H + \frac{\omega_{,\phi}}{\omega}\dot{\phi}\right)\dot{\varphi} + \left[\frac{k^2}{a^2} + \left(\frac{\omega_{,\phi}}{\omega}\right)_{,\phi}\frac{\dot{\phi}^2}{2} + \left(\frac{2U_{,\phi} - F_{,\phi}R}{2\omega}\right)_{,\phi}\right]\varphi \\
 = \dot{\phi}\dot{\Psi} + \left(2\ddot{\phi} + 3H\dot{\phi} + \frac{\omega_{,\phi}}{\omega}\dot{\phi}^2\right)\Psi + 3\dot{\phi}(H\Psi - \dot{\Phi}) + \frac{1}{2\omega}F_{,\phi}\delta R, \quad (11.158)
 \end{aligned}$$

$$\Psi + \Phi = -\frac{\delta F}{F} = -\frac{F_{,\phi}}{F}\varphi, \quad (11.159)$$

where $\varphi = \delta\phi$ is the perturbed field, $\omega = (1 - 6Q^2)F$, and

$$\delta R = 2\left[3(\dot{\Phi} - H\Psi) - 12H(H\Psi - \dot{\Phi}) + \left(\frac{k^2}{a^2} - 3\dot{H}\right)\Psi + 2\frac{k^2}{a^2}\Phi\right]. \quad (11.160)$$

As long as the mass M defined in Eq. (11.156) is sufficiently heavy to satisfy the conditions $M^2 \gg R$, we can approximate $[(2U_{,\phi} - F_{,\phi}R)/2\omega]_{,\phi} \simeq M^2/\omega$ in Eq. (11.158). The solution to Eq. (11.158) consists of the sum of the matter-induced mode φ_{ind} sourced by the matter perturbation and the oscillating mode φ_{osc} , i.e. $\varphi = \varphi_{\text{ind}} + \varphi_{\text{osc}}$. The oscillating mode corresponds to the solution of Eq. (11.158) without the matter perturbation.

Let us first derive the matter-induced mode on sub-horizon scales. In so doing we use the approximation that the terms containing k^2/a^2 , $\delta\rho_m$, δR , and M^2 are the dominant contributions in Eqs. (11.157)–(11.160).¹ Under this approximation, we have $\delta R_{\text{ind}} \simeq 2(k^2/a^2)[\Phi - (F_{,\phi}/F)\varphi_{\text{ind}}]$ from Eqs. (11.159) and (11.160), where the subscript “ind” represents the matter-induced mode. Then from Eq. (11.158)

¹ This approximation was first used in Ref. [374] for the scalar-tensor theory with the Lagrangian density $\mathcal{L} = (1/2)F(\phi)R - (1/2)\nabla(\phi)^2 - U(\phi)$ in the massless limit: $M^2 \ll k^2/a^2$.

we find

$$\varphi_{\text{ind}} \simeq -\frac{2QF}{(k^2/a^2)(1-2Q^2)F + M^2} \frac{k^2}{a^2} \Phi. \quad (11.161)$$

Using Eqs. (11.157) and (11.159) we obtain

$$\begin{aligned} \frac{k^2}{a^2} \Psi &\simeq -\frac{\delta\rho_m}{2F} \frac{(k^2/a^2)(1+2Q^2)F + M^2}{(k^2/a^2)F + M^2}, \\ \frac{k^2}{a^2} \Phi &\simeq \frac{\delta\rho_m}{2F} \frac{(k^2/a^2)(1-2Q^2)F + M^2}{(k^2/a^2)F + M^2}. \end{aligned} \quad (11.162)$$

In the massive limit $M^2/F \gg k^2/a^2$, we recover the standard result of General Relativity. In the massless limit $M^2/F \ll k^2/a^2$, one has $(k^2/a^2)\Psi \simeq -(\delta\rho_m/2F)(1+2Q^2)$ and $(k^2/a^2)\Phi \simeq (\delta\rho_m/2F)(1-2Q^2)$. Note that this recovers Eq. (11.148) in $f(R)$ gravity by setting $Q = -1/\sqrt{6}$.

Substituting Eq. (11.162) into Eq. (11.131), we obtain the equation for matter perturbations

$$\delta_m'' + \left(1 + \frac{\mathcal{H}'}{\mathcal{H}}\right) \delta_m' - \frac{3}{2} \Omega_m \delta_m \frac{(k^2/a^2)(1+2Q^2)F + M^2}{(k^2/a^2)F + M^2} = 0. \quad (11.163)$$

This equation can be also written as [273]

$$\delta_m'' + \left(1 + \frac{\mathcal{H}'}{\mathcal{H}}\right) \delta_m' - \frac{4\pi G_{\text{eff}} \rho_m \delta_m}{H^2} = 0, \quad (11.164)$$

where the (cosmological) effective gravitational “constant” is

$$G_{\text{eff}} = \frac{G}{F} \frac{(k^2/a^2)(1+2Q^2)F + M^2}{(k^2/a^2)F + M^2} = \frac{G}{F} \left[1 + \frac{2Q^2 k^2}{k^2 + a^2 M^2/F}\right]. \quad (11.165)$$

Note that we have recovered the bare gravitational constant G by using $G = 1/(8\pi)$. In the massless limit this reduces to

$$G_{\text{eff}} \simeq \frac{G}{F} (1+2Q^2) = \frac{G}{F} \frac{4+2\omega_{\text{BD}}}{3+2\omega_{\text{BD}}} \quad (M^2/F \ll k^2/a^2), \quad (11.166)$$

where in the last line we have used the relation (9.122) between the coupling Q and the Brans–Dicke parameter ω_{BD} . Note that the cosmological effective gravitational constant (11.166) agrees with the Newton gravitational constant (9.154).

Let us pause to compare the Jordan frame G_{eff} in Eq. (11.165) with the similar result (11.113) in the Einstein frame. The two expressions are in a similar, but not identical, form if we identify $M/F^{1/2}$ as the field dimensionless mass m . The difference is due to the fact that the real observable is not G_{eff} but rather the dimensionless (if $c = 1$) combination $G_{\text{eff}} M/r$: it is this combination that gives the strength of the interaction. Upon a conformal transformation with conformal factor

$F = e^{-2Q\phi}$ as in Section 9.1.3 the distances in the two frames scale as $\tilde{r} = F^{1/2}r$ (as usual the tilded quantity is in the Einstein frame). In coupled dark energy the physical frame is the Einstein one and we have from (11.117):

$$\frac{G_{\text{eff}}M(\phi)}{\tilde{r}} = \frac{GM_0e^{Q\phi}(1+Y_c)}{e^{-Q\phi}r} = \frac{GM_0(1+Y_c)}{re^{-2Q\phi}}, \quad (11.167)$$

where Y_c is the Yukawa correction. It appears therefore that this coincides indeed with the Jordan frame result

$$\frac{G_{\text{eff}}M_0}{r} = \frac{GM_0(1+Y_c)}{Fr} = \frac{GM_0(1+Y_c)}{re^{-2Q\phi}}. \quad (11.168)$$

Next, let us derive the approximate equation for the oscillating mode. Using Eqs. (11.157) and (11.158) under the condition $k^2/a^2 \gg H^2$ the gravitational potentials for $\delta\rho_m = 0$ are expressed by φ_{osc} . Then from Eq. (11.160) the perturbation δR corresponding to the oscillating mode is given by

$$\delta R_{\text{osc}} \simeq 6Q \left(\ddot{\varphi}_{\text{osc}} + 3H\dot{\varphi}_{\text{osc}} + \frac{k^2}{a^2}\varphi_{\text{osc}} \right). \quad (11.169)$$

Substituting this relation into Eq. (11.158), we find

$$\ddot{\varphi}_{\text{osc}} + 3H\dot{\varphi}_{\text{osc}} + \left(\frac{k^2}{a^2} + \frac{M^2}{F} \right) \varphi_{\text{osc}} \simeq 0, \quad (11.170)$$

which is valid in the regime $M^2 \gg R$.

When $|Q| = \mathcal{O}(1)$ the field potential $U(\phi)$ needs to be heavy in the region of high density for the consistency with local gravity constraints. We shall consider the potential (9.148) as an example of a viable model. During the matter era the field ϕ sits at the instantaneous minima characterized by the condition (9.150). Hence, we have the relations $\phi \propto \rho_m^{\frac{1}{p-1}}$ and $M^2 \propto \rho_m^{\frac{2-p}{1-p}}$ during the matter-dominated epoch. The field ϕ can initially be heavy to satisfy the condition $M^2/F \gg k^2/a^2$ for the modes relevant to the galaxy power spectrum. Depending upon the model parameters and the mode k , the mass squared M^2 can be smaller than k^2/a^2 during the matter era [273].

In the regime $M^2/F \gg k^2/a^2$ the matter perturbation equation (11.163) reduces to the standard one in Einstein gravity, which gives the evolution $\delta_m \propto t^{2/3}$. For the model (9.148) the matter-induced mode of the field perturbation evolves as $\varphi_{\text{ind}} \propto \delta\rho_m/M^2 \propto t^{\frac{2(4-p)}{3(1-p)}}$. Meanwhile, the WKB solution to Eq. (11.170) is given by $\varphi_{\text{osc}} \propto t^{\frac{p}{2(1-p)}} \cos(ct^{-\frac{1}{1-p}})$, where c is a constant. Since the background field ϕ during the matter era evolves as $\phi \propto t^{\frac{2}{1-p}}$, we find

$$\varphi/\phi = (\varphi_{\text{ind}} + \varphi_{\text{osc}})/\phi \simeq c_1 t^{2/3} + c_2 t^{-\frac{4-p}{2(1-p)}} \cos\left(ct^{-\frac{1}{1-p}}\right). \quad (11.171)$$

As long as the oscillating mode is initially suppressed relative to the matter-induced mode, the matter-induced mode remains the dominant contribution. This property also holds during the radiation-dominated epoch.

In the regime $M^2/F \ll k^2/a^2$ the effective gravitational constant is given by Eq. (11.166), which shows that the effect of modified gravity becomes important. Solving Eq. (11.164) in this case, we obtain the solution for matter perturbations

$$\delta_m \propto t^{\frac{\sqrt{25+48Q^2}-1}{6}}. \quad (11.172)$$

Setting $Q = -1/\sqrt{6}$, this recovers the solution $\delta_m \propto t^{(\sqrt{33}-1)/6}$ in $f(R)$ gravity.

The potential (9.148) has a heavy mass M which is much larger than H in the deep matter-dominated epoch, but it gradually decreases to become of the order of H around the present epoch. Depending on the modes k , the system crosses the point $M^2/F = k^2/a^2$ at $t = t_k$. Since for the potential (9.148) M evolves as $M \propto t^{-\frac{2-p}{1-p}}$ during the matter era, the time t_k has a scale-dependence given by $t_k \propto k^{-\frac{3(1-p)}{4-p}}$. When $t < t_k$ the evolution of δ_m is given by $\delta_m \propto t^{2/3}$, but for $t > t_k$ its evolution changes to the form given by (11.172).

During the matter era the mass squared is approximately given by

$$M^2 \simeq \frac{1-p}{(2^p p C)^{1/(1-p)}} Q^2 \left(\frac{\rho_m}{U_0} \right)^{\frac{2-p}{1-p}} U_0. \quad (11.173)$$

Using the relation $\rho_m = 3F_0\Omega_m^{(0)}H_0^2(1+z)^3$, we find that the critical redshift z_k at time t_k can be estimated as

$$z_k \simeq \left[\left(\frac{k}{a_0 H_0} \frac{1}{|Q|} \right)^{2(1-p)} \frac{2^p p C}{(1-p)^{1-p}} \frac{1}{(3F_0\Omega_m^{(0)})^{2-p}} \frac{U_0}{H_0^2} \right]^{\frac{1}{4-p}} - 1, \quad (11.174)$$

where a_0 is the present scale factor. The critical redshift increases for larger $k/(a_0 H_0)$. The matter power spectrum, in the linear regime, has been observed for the scales $0.01h \text{ Mpc}^{-1} \lesssim k \lesssim 0.2h \text{ Mpc}^{-1}$, which corresponds to $30a_0 H_0 \lesssim k \lesssim 600a_0 H_0$. In Fig. 11.4 we plot the evolution of the growth rate $f = \dot{\delta}_m/(H\delta_m)$ for the mode $k = 600a_0 H_0$ and the coupling $Q = 1.08$ with three different values of p . Note that the asymptotic values of f in the regions $t \ll t_k$ and $t \gg t_k$ are given by $f = 1$ and $f = (\sqrt{25+48Q^2}-1)/4$, respectively. We find that, for the scales $30a_0 H_0 \lesssim k \lesssim 600a_0 H_0$, the critical redshift exists in the region $z_k \gtrsim 1$ and that z_k increases for smaller p . When $p = 0.7$ we have $z_k = 3.9$ from Eq. (11.174), which is consistent with the numerical result in Fig. 11.4. The growth rate f reaches a maximum value $f_{\max}$ and then begins to decrease around the end of the matter era.

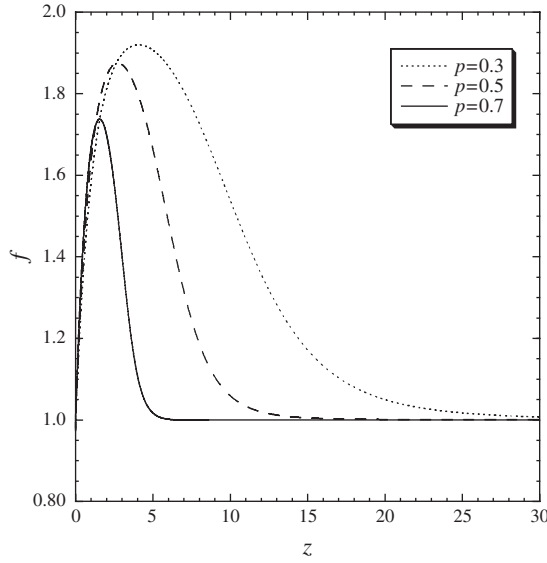


Figure 11.4 The evolution of the growth rate f of matter perturbations in terms of the redshift z for $Q = 1.08$ and $k = 600a_0H_0$ with three different values of p . For smaller p the critical redshift z_k gets larger. The growth rate f reaches a maximum value and begins to decrease after the system enters the accelerated epoch. For smaller p the maximum value of f tends to approach the analytic value $f = (\sqrt{25 + 48Q^2} - 1)/4$. From Ref. [273].

McDonald *et al.* [493] derived the constraint $f = 1.46 \pm 0.49$ around the redshift, $z = 3$, from the measurement of the matter power spectrum from the Lyman- α forests. The more recent data reported by Viel and Haehnelt [494] in the redshift range $2 < z < 4$ show that even the value $f = 2$ can be allowed in some of the observations. If we use the criterion $f < 2$ for the analytic estimation $f = (\sqrt{25 + 48Q^2} - 1)/4$, we obtain the bound $Q < 1.08$. Figure 11.4 shows that $f_{\max}$ is smaller than the analytic value $f = 2$ (which corresponds to $Q = 1.08$). When $p = 0.7$, for example, we have that $f_{\max} = 1.74$. For the values of p that are very close to 1, $f_{\max}$ can be smaller than 1.5. However these cases are hardly distinguishable from the Λ CDM model. In any case the current observational data on the growth rate f are not enough to place tight bounds on Q and p .

As in the case of $f(R)$ gravity, the matter power spectrum P_{δ_m} at time $t = t_\Lambda$ (at which $\ddot{a} = 0$) shows a difference compared to the Λ CDM model given by

$$\frac{P_{\delta_m}(t_\Lambda)}{P_{\delta_m}^{\Lambda\text{CDM}}(t_\Lambda)} = \left(\frac{t_\Lambda}{t_k}\right)^{2\left(\frac{\sqrt{25+48Q^2}-1}{6}-\frac{2}{3}\right)} \propto k^{\frac{(1-p)(\sqrt{25+48Q^2}-5)}{4-p}}, \quad (11.175)$$

for small ($k \gg MaF^{-1/2}$), but still linear, scales. The CMB power spectrum is also modified by the non-standard evolution of the effective gravitational potential

$$\psi = \Phi - \Psi = \frac{3a^2 H^2}{k^2} \Omega_m \delta_m, \quad (11.176)$$

which mainly affects the low multipoles because of the ISW effect. Since the smaller scale modes in CMB relevant to the galaxy power spectrum are hardly affected by this modification, there is a difference between the spectral indices of the matter power spectrum and of the CMB spectrum on the scales, $k > 0.01h \text{ Mpc}^{-1}$:

$$\Delta n(t_\Lambda) = \frac{(1-p)(\sqrt{25+48Q^2}-5)}{4-p}. \quad (11.177)$$

This reproduces the result (11.154) in $f(R)$ gravity by setting $Q = -1/\sqrt{6}$ and $p = 2n/(2n+1)$. If we use the criterion $\Delta n(t_\Lambda) < 0.05$, as in the case of the $f(R)$ gravity, we obtain the bounds $p > 0.96$ for $Q = 1$ and $p > 0.86$ for $Q = 0.5$. As long as p is close to 1, it is possible to satisfy both cosmological and local gravity constraints for $|Q| \lesssim 1$.

11.6.3 DGP braneworld model

Finally, we discuss the evolution of linear matter perturbations in the DGP braneworld model. The perturbed metric in the 5-dimensional longitudinal gauge with four scalar metric perturbations Ψ, Φ, B, E is given by [441, 495]

$$\begin{aligned} ds^2 = & -(1+2\Psi)n(t, y)^2 dt^2 + (1+2\Phi)A(t, y)^2 \delta_{ij} dx^i dx^j \\ & + 2r_c B_{,i} dx^i dy + (1+2E)dy^2, \end{aligned} \quad (11.178)$$

where the brane is located at $y = 0$ in the 5-th dimension characterized by the coordinate y (we are considering a flat FLRW spacetime on the brane). Note that B can be identified as a brane bending mode describing a perturbation of the brane location and that r_c is the crossover scale defined in Eq. (9.215). The solution for the background metric describing the self-accelerating Universe is [433]

$$n(t, y) = 1 + H(1 + \dot{H}/H^2)y, \quad A(t, y) = a(t)(1 + Hy). \quad (11.179)$$

Recall that the Hubble parameter $H = \dot{a}/a$ satisfies Eq. (9.218) with $\epsilon = +1$.

In the following we shall neglect the terms suppressed by the factor $aH/k \ll 1$ because we are considering sub-horizon perturbations. We also neglect the terms such as $(A'/A)\Phi'$, where a prime represents a derivative with respect to y . This comes from the fact that Φ' is of the order of $(k/a)\Phi$, as we will show later. The time-derivative terms can be also dropped under a quasi-static approximation. Then

the perturbed 5-dimensional Einstein tensors $\delta\tilde{G}_B^A$ obey the following equations locally in the bulk [495]:

$$\delta\tilde{G}_0^0 = 3\Phi'' + \frac{2}{A^2}\nabla^2\Phi + \frac{\nabla^2}{A^2}(E - r_c B') - 2\frac{r_c}{A^2}\left(\frac{A'}{A}\right)\nabla^2 B = 0, \quad (11.180)$$

$$\begin{aligned} \delta\tilde{G}_j^i = & -\frac{1}{A^2}(\nabla^i\nabla_j - \delta_j^i\nabla^2)(\Phi + \Psi + E - r_c B') + \delta_j^i(\Psi'' + 2\Phi'') \\ & + \frac{r_c}{A^2}(\nabla^i\nabla_j - \delta_j^i\nabla^2)\left(\frac{A'}{A} + \frac{n'}{n}\right)B = 0, \end{aligned} \quad (11.181)$$

$$\delta\tilde{G}_i^5 = -(\Psi' + 2\Phi')_{,i} = 0, \quad (11.182)$$

$$\delta\tilde{G}_5^5 = \frac{1}{A^2}\nabla^2(\Psi + 2\Phi) - \frac{r_c}{A^2}\left(2\frac{A'}{A} + \frac{n'}{n}\right)\nabla^2 B = 0. \quad (11.183)$$

Taking the divergence of the traceless part of Eq. (11.181), we get

$$\frac{\nabla^2}{A^2}(\Phi + \Psi + E - r_c B') - \frac{r_c}{A^2}\left(\frac{A'}{A} + \frac{n'}{n}\right)\nabla^2 B = 0. \quad (11.184)$$

For the consistency between Eqs. (11.182) and (11.183), it is required that

$$B' = 0, \quad \Psi' + 2\Phi' = 0. \quad (11.185)$$

From Eqs. (11.183) and (11.184) we obtain

$$\frac{\nabla^2}{A^2}(E - r_c B') = -\frac{1}{2}\frac{\nabla^2}{A^2}\Psi + \frac{r_c}{2A^2}\frac{n'}{n}\nabla^2 B. \quad (11.186)$$

Substituting Eqs. (11.183) and (11.186) into Eq. (11.180) together with the use of Eq. (11.185), we find

$$\Psi'' + \frac{\nabla^2}{A^2}\Psi - \frac{n'}{n}\frac{r_c}{A^2}\nabla^2 B = 0. \quad (11.187)$$

Under the sub-horizon approximation ($k \gg aH$) the solution of Eq. (11.187), upon the Fourier transformation, is given by

$$\Psi - \frac{n'}{n}r_c B = [c_1(1 + Hy)^{-k/aH} + c_2(1 + Hy)^{k/aH}], \quad (11.188)$$

where c_1 and c_2 are integration constants. In order to avoid the divergence of the perturbation in the limit $y \rightarrow \infty$ we shall choose $c_2 = 0$.

The junction condition at the brane can be written in terms of an extrinsic curvature $K_{\mu\nu}$ and an energy-momentum tensor on the brane:

$$K_{\mu\nu} - Kg_{\mu\nu} = -\frac{\kappa_{(5)}^2}{2}T_{\mu\nu} + r_c G_{\mu\nu}, \quad (11.189)$$

where $K \equiv K_\mu^\mu$. Note that the extrinsic curvature is defined as $K_{\mu\nu} = h_\mu^\lambda \nabla_\lambda n_\nu$, where n_ν is the unit vector normal to the brane and $h_{\mu\nu} = g_{\mu\nu} - n_\mu n_\nu$ is the induced metric on the brane. The (0, 0) and spatial components of the junction condition (11.189) give

$$\frac{2}{a^2} \nabla^2 \Phi = -\kappa_{(4)}^2 \delta \rho_m + \frac{1}{a^2} \nabla^2 B - \frac{3}{r_c} \Phi', \quad (11.190)$$

$$\Phi + \Psi = B, \quad (11.191)$$

$$\Psi' + 2\Phi' = 0, \quad (11.192)$$

where $\delta \rho_m$ is the matter perturbation on the brane. Note that Eq. (11.192) is consistent with the latter of Eq. (11.185).

From Eq. (11.188) it follows that $\Phi' \sim (k/a)\Phi$ in Fourier space. For the perturbations whose wavelengths are much smaller than the cross-over scale r_c , i.e. $r_c k/a \gg 1$, we find that the term $(3/r_c)\Phi'$ in Eq. (11.190) is much smaller than $(k^2/a^2)\Phi$. In Fourier space Eq. (11.190) is approximately given by

$$\frac{2k^2}{a^2} \Phi = \kappa_{(4)}^2 \delta \rho_m + \frac{k^2}{a^2} B. \quad (11.193)$$

Using the projection of Eq. (11.183) as well as Eqs. (11.191) and (11.193), we find that metric perturbations Ψ and Φ obey the following equations

$$\frac{k^2}{a^2} \Psi = -\frac{\kappa_{(4)}^2}{2} \left(1 + \frac{1}{3\beta}\right) \delta \rho_m, \quad \frac{k^2}{a^2} \Phi = \frac{\kappa_{(4)}^2}{2} \left(1 - \frac{1}{3\beta}\right) \delta \rho_m, \quad (11.194)$$

where

$$\beta(t) \equiv 1 - \frac{2r_c}{3} \left(2\frac{A'}{A} + \frac{n'}{n}\right) = 1 - 2Hr_c \left(1 + \frac{\dot{H}}{3H^2}\right). \quad (11.195)$$

The matter perturbation δ_m satisfies the same form of equation as given in (11.131) for the modes deep inside the horizon [440, 441]. Substituting the former of Eq. (11.194) into Eq. (11.131), we find that the matter perturbation obeys the following equation

$$\delta_m'' + \left(1 + \frac{\mathcal{H}'}{\mathcal{H}}\right) \delta_m' - \frac{3}{2} \left(1 + \frac{1}{3\beta}\right) \Omega_m \delta_m = 0, \quad (11.196)$$

where $\Omega_m \equiv \kappa_{(4)}^2 \rho_m / (3H^2)$.

In the deep matter era one has $Hr_c \gg 1$ and hence $\beta \simeq -Hr_c$, so that β is largely negative ($|\beta| \gg 1$). In this regime the evolution of the matter perturbation is similar to that in General Relativity ($\delta_m \propto t^{2/3}$). The system finally approaches the de Sitter solution characterized by $H_{\text{ds}} = 1/r_c$. We then have $\beta \simeq 1 - 2Hr_c \simeq -1$ around the de Sitter solution. Since $1 + 1/(3\beta) \simeq 2/3$, the growth rate in this regime is

smaller relative to the case of General Relativity. The index γ of the growth rate $f = \Omega_m^\gamma$ is approximated by $\gamma \approx 0.68$ [480], which is different from the value $\gamma \simeq 0.55$ for the Λ CDM model. If the future imaging survey of galaxies can constrain γ within 20%, it may be possible to distinguish the Λ CDM model from DGP modified gravity observationally [496, 497].

From Eq. (11.196) one can regard that the effective gravitational “constant” is given by $G_{\text{eff}} = G[1 + 1/(3\beta)]$. Comparing this with the effective gravitational constant (11.166) in Brans–Dicke theory with a massless limit (or the absence of the field potential), we find that the Brans–Dicke parameter ω_{BD} has the following relation with β :

$$\omega_{\text{BD}} = \frac{3}{2}(\beta - 1). \quad (11.197)$$

Since $\beta < 0$ for the self-accelerating DGP solution, this implies that $\omega_{\text{BD}} < -3/2$. This corresponds to the theory with ghosts, because the kinetic energy of a scalar field degree of freedom becomes negative in the Einstein frame [404]. The DGP ghost is a ghost mode in the scalar sector of the gravitational field, which is more serious than the ghost in a phantom scalar field. Note that another normal branch of solutions in the DGP model does not suffer from this problem because the minus sign of Eq. (11.195) is replaced by the plus sign. In other words, the self-accelerating solution in the original DGP model can be realized at the expense of an appearance of the ghost state.

11.6.4 Reconstruction of the metric at first order

We have shown that modified gravity models generally lead to a change of the growth rate of matter perturbations relative to the Λ CDM model. Given the matter perturbation δ_m , the first-order metric perturbations Φ, Ψ are completely fixed by solving the conservation equations. Since there are two free functions that determine the (scalar part of the) first-order metrics Ψ and Φ , dark energy models can be classified according to how the gravitational potentials are linked to δ_m . In order to quantify this, we introduce two quantities $q(k, t)$ and $\zeta(k, t)$ defined by

$$\frac{k^2}{a^2}\Phi = 4\pi Gq\delta_m\rho_m, \quad (11.198)$$

$$\frac{\Phi + \Psi}{\Phi} = \zeta, \quad (11.199)$$

where G is the 4-dimensional bare gravitational constant. Note that ζ characterizes the strength of the anisotropic stress. The Λ CDM model corresponds to $q = 1$ and $\zeta = 0$ (recall that the cosmological constant does not cluster). A non-clustering

dark energy in Einstein gravity will also have $q = 1, \zeta = 0$ at small scales. In this case the background evolution fixes the equation of state of dark energy and from this one can derive all the cosmological dynamics. On the other hand, any model in which dark energy clusters or gravity is modified (or both) can induce different values of q, ζ relative to the Λ CDM model. Therefore the functions q and ζ completely characterize a gravity theory for first-order scalar perturbations on small scales.

For instance, if gravity is Einsteinian but $\delta_{\text{DE}} \neq 0$, we find from Eq. (11.41) that the gravitational potential Φ on sub-horizon scales satisfies

$$\frac{k^2}{a^2} \Phi = 4\pi G \left(1 + \frac{\Omega_{\text{DE}} \delta_{\text{DE}}}{\Omega_m \delta_m} \right) \delta_m \rho_m, \quad (11.200)$$

which gives $q = 1 + \Omega_{\text{DE}} \delta_{\text{DE}} / (\Omega_m \delta_m)$. Hence the clustered dark energy leads to the deviation from the Λ CDM model.

In the scalar-tensor model discussed in Section 11.6.2, the gravitational potentials are given by Eq. (11.162) on sub-horizon scales. In this case we have

$$q = \frac{1}{F} \frac{(k^2/a^2)(1 - 2Q^2)F + M^2}{(k^2/a^2)F + M^2}, \quad \zeta = -\frac{4F(k^2/a^2)Q^2}{(k^2/a^2)(1 - 2Q^2)F + M^2}, \quad (11.201)$$

where we have used the unit $8\pi G = 1$. In the regime $M^2/F \gg k^2/a^2$ (and $F \simeq 1$) it follows that $q \simeq 1$ and $\zeta \simeq 0$. In the regime $M^2/F \ll k^2/a^2$ we have $q \simeq (1 - 2Q^2)/F$ and $\zeta \simeq -4Q^2/(1 - 2Q^2)$, so that the deviation from the Λ CDM model becomes important. Recall that the expression (11.201) covers the case of $f(R)$ gravity by setting $Q = -1/\sqrt{6}$. In Ref. [498] the quantities q and ζ have been evaluated for the more general Lagrangian density $f(R, \phi, X)$.

In the DGP model the gravitational potentials obey Eq. (11.194), which gives

$$q = 1 - \frac{1}{3\beta}, \quad \zeta = \frac{2}{1 - 3\beta}. \quad (11.202)$$

In the deep matter era one has $|\beta| \gg 1$, so that $q \simeq 1$ and $\zeta \simeq 0$. The deviation from $(q, \zeta) = (1, 0)$ appears when $|\beta|$ decreases to the order of unity, i.e. when the Universe enters the epoch of late-time cosmic acceleration.

In order to confront dark energy models with the observations of weak lensing, it may be convenient to introduce the following quantity [497]

$$\Sigma \equiv q(1 - \zeta/2). \quad (11.203)$$

From Eqs. (11.198) and (11.199) we find that the weak lensing potential $\psi = \Phi - \Psi$ can be expressed as

$$\psi = 8\pi G \frac{a^2}{k^2} \rho_m \delta_m \Sigma . \quad (11.204)$$

Notice that for the DGP model we have $\Sigma = 1$ while for scalar-tensor models $\Sigma = 1/F$. This is ultimately a consequence of the fact that scalar-tensor theories are conformally equivalent to ordinary gravity and therefore null geodesics remain the same (i.e. the equation $ds^2 = 0$ is conformally invariant). Then the photon propagation equation (4.213) and its first-order version (4.222) are unchanged and thus the lensing potential ψ is unchanged as well, except for an overall rescaling.

The effect of modified gravity theories manifests itself in weak lensing observations in at least two ways. One is the multiplication of the term Σ on the r.h.s. of Eq. (11.204). Another is the modification of the evolution of δ_m . The latter depends on the two parameters q and ζ , or, equivalently, Σ and ζ . These two parameters (Σ, ζ) will be useful to detect signatures of modified gravity theories from future surveys of weak lensing.

11.7 Problems

- 11.1** Take Eq. (11.43) and find the analytical solution in the case where Ω_{DE} and w are constants.
- 11.2** In $f(R)$ gravity, find the (cosmological) effective gravitational “constant” G_{eff} from the equation of matter perturbations.

Non-linear cosmological perturbations

Cosmology is, by and large, the realm of linear gravitational processes. When gravitational instability reaches a regime of non-linearity, astrophysical objects form (galaxies, black hole, stars) and the memory of the global structure of spacetime, and therefore of cosmology, is lost or diluted in new physics and new interactions. There is however an intermediate regime in which gravity is still the only player but effects beyond linearity begin to be observable. This regime lies between the linear perturbation theory we have explored in the previous chapter and the full non-linear dynamics that can be dealt with only in N -body simulations or by focusing on single objects.

It may be expected that the presence of dark energy will not influence small-scale non-linear processes. This is probably true for standard dark energy, i.e. a smooth component with negligible clustering described by a slowly varying equation of state. However we have learned how rich the possible phenomenology of dark energy is. We cannot exclude that weakly non-linear processes might keep some record on the cosmological conditions in which they developed. We know of at least one such process, the epoch of the beginning of structure formation and, as a consequence, the abundance of collapsed objects.

This chapter will present the effects of non-linearity in higher-order cosmological perturbation theory that are of interest in dark energy research. Primes here denote differentiation with respect to $N = \ln a$.

12.1 Second-order perturbations

We have seen that the perturbation equations in the sub-horizon regime reproduce the linearized, Newtonian versions of the continuity equation, the Euler equation, and the Poisson equation of classical fluid dynamics. We could then as well make a step further and derive the full equation, without linearization. In fact, at least for dust, they remain identical to the Newtonian laws in physical coordinates.

Consider a pressureless perfect fluid with density ρ moving with a velocity $\mathbf{u}$ under an influence of gravity characterized by a gravitational potential Φ_N . The continuity equation (the conservation of mass) and the Euler equation (the Newton's equation of motion with the gravitational force $\mathbf{f} = -\nabla_r \Phi_N$) in the fluid dynamics are given, respectively, by

$$\left(\frac{\partial \rho}{\partial t} \right)_r + \nabla_r \cdot (\rho \mathbf{u}) = 0, \quad (12.1)$$

$$\left(\frac{\partial \mathbf{u}}{\partial t} \right)_r + (\mathbf{u} \cdot \nabla_r) \mathbf{u} = -\nabla_r \Phi_N, \quad (12.2)$$

where Φ_N is the gravitational potential in Newton's gravity satisfying the Poisson equation

$$\nabla_r^2 \Phi_N = 4\pi G \rho. \quad (12.3)$$

The subscript r means that the physical (proper) coordinate $\mathbf{r}$ is used in Eqs. (12.1) and (12.2).

In the expanding Universe it is convenient to change variables from the proper locally Minkowski coordinate $\mathbf{r}$ to the expanding coordinate $\mathbf{x} = \mathbf{r}/a(t)$ comoving in the background model. Then the fluid velocity $\mathbf{u} = \dot{\mathbf{r}}$ is given by

$$\mathbf{u} = \dot{a}\mathbf{x} + \mathbf{v}(\mathbf{x}, t), \quad (12.4)$$

where $\mathbf{v} = a\dot{\mathbf{x}}$ is the peculiar velocity. Note that the transformation laws hold

$$\nabla_x = a \nabla_r, \quad \left(\frac{\partial f(\mathbf{x}, t)}{\partial t} \right)_r = \left(\frac{\partial f}{\partial t} \right)_x - \frac{\dot{a}}{a} (\mathbf{x} \cdot \nabla_x) f, \quad (12.5)$$

where f is an arbitrary function with respect to $\mathbf{x}$ and t . We define as usual the matter density contrast

$$\delta(\mathbf{x}, t) \equiv \frac{\delta\rho(\mathbf{x}, t)}{\bar{\rho}(t)}, \quad \delta\rho \equiv \rho(\mathbf{x}, t) - \bar{\rho}(t), \quad (12.6)$$

where $\bar{\rho}(t)$ is the background matter density.

Carrying out the transformation of Eqs. (12.1), (12.2), and (12.3) to the comoving coordinate, we obtain [95, 499]

$$\frac{\partial \delta}{\partial t} + \frac{1}{a} \nabla_x \cdot (1 + \delta) \mathbf{v} = 0, \quad (12.7)$$

$$\frac{\partial \mathbf{v}}{\partial t} + H \mathbf{v} + \frac{1}{a} \mathbf{v} \cdot \nabla_x \mathbf{v} = \frac{1}{a} \nabla_x \Phi, \quad (12.8)$$

$$\frac{1}{a^2} \nabla_x^2 \Phi = -4\pi G \delta\rho, \quad (12.9)$$

where

$$\Phi \equiv - \left(\Phi_N + \frac{1}{2} a \ddot{a} x^2 \right). \quad (12.10)$$

See problem 12.1 for the derivation of Eqs. (12.7)–(12.9). Note that the conservation equation, $\dot{\bar{\rho}}(t) + 3H\bar{\rho}(t) = 0$, also follows from Eq. (12.1) as the zero-th order. In order to derive Eq. (12.9) together with the effective gravitational potential (12.10), we have used the relation $\ddot{a}/a = -(4\pi G/3)\bar{\rho}(t)$. Equations (12.7)–(12.9) are the master equations to describe the evolution of pressureless matter perturbations including the effect of non-linearity. In the following we drop the subscript x for brevity.

For convenience we rewrite Eqs. (12.7)–(12.9) by using the derivative with respect to $N = \ln a$, the velocity vector v_i , and the density parameter $\Omega_m = 8\pi G\rho_m/(3H^2)$:

$$\mathcal{H}\delta' = -\nabla^i[(1+\delta)v_i], \quad (12.11)$$

$$\mathcal{H}v'_i = -\mathcal{H}v_i - v_j \nabla^j v_i + \nabla_i \Phi, \quad (12.12)$$

$$\nabla^2 \Phi = -\frac{3}{2} \Omega_m \mathcal{H}^2 \delta, \quad (12.13)$$

where repeated indices are summed over. All the perturbation variables depend both on space and time.

These equations are derived under a number of simplifying assumptions. First, they refer to pressureless matter ($w = c_s^2 = 0$). Second, they are valid in the sub-horizon, small velocity regime. Since the non-linear effects are expected to be more important at small scales this seems well justified. Third, we have neglected any shear term in the Euler equation, that is, we have assumed that the pressureless fluid has no viscosity. This property allows the set of equations to be closed. The absence of shear implies that spherical perturbations at the initial stage remain spherical throughout.

The velocity vector v_i can be generally decomposed in transverse and parallel parts. As before we neglect the transverse (or vorticity) part on the grounds that if it is initially zero it remains zero throughout – but again for this to be true a vanishing shear stress is required. Building a total derivative

$$\frac{d\delta}{dN} \equiv \delta' + \frac{v^i}{\mathcal{H}} \nabla_i \delta, \quad (12.14)$$

where a prime represents a partial derivative with respect to N , the continuity equation (12.11) can be written as

$$\frac{d\delta}{dN} = -\theta(1+\delta). \quad (12.15)$$

As in the previous chapter, $\theta = \nabla^i v_i / \mathcal{H}$. Taking the divergence of the Euler equation (12.12) and using Eq. (12.13), we obtain

$$\theta' = - \left(1 + \frac{\mathcal{H}'}{\mathcal{H}} \right) \theta - \frac{1}{\mathcal{H}^2} \nabla^i (v_j \nabla^j v_i) - \frac{3}{2} \Omega_m \delta. \quad (12.16)$$

Now we have

$$\nabla^i (v_j \nabla^j v_i) = (\nabla^i v_j)(\nabla^j v_i) + v_j \nabla^j \nabla^i v_i. \quad (12.17)$$

The last term is $\mathcal{H} v_j \nabla^j \theta$ and can be absorbed into the l.h.s. of Eq. (12.16) to form a total derivative $d\theta/dN$. The first term on the r.h.s. of Eq. (12.17) can be simplified by recalling that the shear-free peculiar velocity field in an initially spherical perturbation remains purely radial and therefore depends only on r . In other words, one can write $\mathbf{v} = (v/\sqrt{3})\{1, 1, 1\}$ so that

$$(\nabla^i v_j)(\nabla^j v_i) = \frac{1}{3} \mathcal{H}^2 \theta^2. \quad (12.18)$$

So Eq. (12.16) depends on θ, δ alone:

$$\frac{d\theta}{dN} = - \left(1 + \frac{\mathcal{H}'}{\mathcal{H}} \right) \theta - \frac{1}{3} \theta^2 - \frac{3}{2} \Omega_m \delta. \quad (12.19)$$

Notice that we do not need to derive the peculiar velocity to second-order since the second-order terms are proportional to the gravitational potential Φ [compare Eq. (4.28)] and are therefore negligible on small scales. From Eqs. (12.15) and (12.19) we obtain

$$\frac{d^2 \delta}{dN^2} + \left(1 + \frac{\mathcal{H}'}{\mathcal{H}} \right) \frac{d\delta}{dN} - \frac{3}{2} \Omega_m \delta = \frac{4}{3} \frac{1}{1 + \delta} \left(\frac{d\delta}{dN} \right)^2 + \frac{3}{2} \Omega_m \delta^2, \quad (12.20)$$

where on the l.h.s. we retrieve the usual linearized terms of Eq. (4.104) while on the r.h.s we have the new non-linear terms. Since we are still assuming $\delta = \delta(\mathbf{x}, a)$ the total derivative includes the spatial part as well.

If, more in general, we do not make the assumption of radial perturbations, Eq. (12.20) becomes

$$\frac{d\delta^2}{dN^2} + \left(1 + \frac{\mathcal{H}'}{\mathcal{H}} \right) \frac{d\delta}{dN} - \frac{3}{2} \Omega_m \delta = \frac{1}{1 + \delta} \left(\frac{d\delta}{dN} \right)^2 + \frac{3}{2} \Omega_m \delta^2 + (\nabla^i v_j)(\nabla^j v_i) \frac{1 + \delta}{\mathcal{H}^2}. \quad (12.21)$$

In order to see the effects of non-linear terms, we expand $\delta(\mathbf{x}, a)$ in a perturbative series, $\delta = \sum_{n=1}^{\infty} \delta^{(n)}$, where the terms are assumed to be of order n . At first order we are then left with the usual separable linear perturbation equation

$\delta^{(1)} = \delta_1(\mathbf{x})D_L(a)$. You can show in problem 12.2 that at second order we obtain

$$\delta^{(2)''} + \left(1 + \frac{\mathcal{H}'}{\mathcal{H}}\right) \delta^{(2)'} - \frac{3}{2}\Omega_m \delta^{(2)} = \frac{3}{2} \frac{\Omega_m}{f^2} D_L'^2 [\delta_1^2 + \delta_{1,i} \Delta_{,i}] + D_L'^2 [\delta_1^2 + 2\delta_{1,i} \Delta_{,i} + \Delta_{,ij} \Delta_{,ij}], \quad (12.22)$$

where $f \equiv D_L'/D_L$ is the linear growth rate, and

$$\Delta(\mathbf{x}) \equiv -\frac{1}{4\pi} \int \frac{\delta_1(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} d^3x'. \quad (12.23)$$

Note that spatial derivatives are denoted with commas and that $\Delta(\mathbf{x})$ is the general solution of the Poisson equation $\nabla^2 \Delta = \delta_1(\mathbf{x})$.

Following Ref. [506] we can now solve Eq. (12.22) by assuming that the solution is dominated by the inhomogeneous terms (which is indeed the case). If we denote with D_{2a} a particular solution of

$$D_{2a}'' + \left(1 + \frac{\mathcal{H}'}{\mathcal{H}}\right) D_{2a}' - \frac{3}{2}\Omega_m D_{2a} = \frac{3}{2} \frac{\Omega_m}{f^2} D_L'^2, \quad (12.24)$$

and with D_{2b} a particular solution of

$$D_{2b}'' + \left(1 + \frac{\mathcal{H}'}{\mathcal{H}}\right) D_{2b}' - \frac{3}{2}\Omega_m D_{2b} = D_L'^2, \quad (12.25)$$

then the general solution can be written as

$$\delta^{(2)} = (D_{2a} + D_{2b})\delta_1^2 + (D_{2a} + 2D_{2b})\delta_{1,i} \Delta_{,i} + D_{2b} \Delta_{,ij} \Delta_{,ij}, \quad (12.26)$$

where $D_{2b} = (D_L'^2 - D_{2a})/2$. To respect the assumption of initial Gaussianity we assume $D_{2a}' = D_{2a} = 0$ at the initial time (i.e. far into the past). For a spherical perturbation, the last term on the r.h.s. in Eq. (12.26) is $D_{2b}\delta_1^2/3$ and sums with the first to give $(D_{2a} + 4D_{2b}/3)\delta_1^2$. If, moreover, the gradient term $\delta_{1,i} \Delta_{,i}$ is negligible (for instance near the center of a top-hat perturbation), then the second-order solution is separable, $\delta^{(2)} = \delta_1^2(\mathbf{x})D_2(a)$, where D_2 is the solution of the equation

$$D_2'' + \left(1 + \frac{\mathcal{H}'}{\mathcal{H}}\right) D_2' - \frac{3}{2}\Omega_m D_2 = \left(\frac{3}{2} \frac{\Omega_m}{f^2} + \frac{4}{3}\right) D_L'^2. \quad (12.27)$$

This can be obtained from Eq. (12.20) by neglecting the gradient terms in the total derivatives. This approximation can be extended to all orders. In the simple Einstein–de Sitter case ($\mathcal{H}'/\mathcal{H} = -1/2$ and $\Omega_m = 1$) all coefficients in Eq. (12.27) are constant and to order $n = 2, 3$ we obtain (you can check by direct

substitution) [95]

$$D_2 = \frac{17}{21} D_L^2, \quad D_3 = \frac{341}{567} D_L^3. \quad (12.28)$$

12.2 The bispectrum and the higher-order correction to the power spectrum

Let us start again from Eqs. (12.11)–(12.13). Since we are assuming the flow to be directed along the gravitational potential gradient, we can write $v_k^i = k^i \alpha$ in Fourier space. Since $\theta_k = i k_i v_k^i / \mathcal{H}$ we obtain $\alpha = -i \theta_k \mathcal{H} k^{-2}$ and hence

$$v_k^i = -\frac{i k^i}{k^2} \mathcal{H} \theta_k. \quad (12.29)$$

Equation (12.11) can be written as

$$\mathcal{H} \delta' + \mathcal{H} \theta = -\nabla_i (\delta v^i). \quad (12.30)$$

Performing the Fourier transformation and using Eq. (12.29), we find¹

$$\begin{aligned} \int d^3 k \mathcal{H} (\delta'_k + \theta_k) e^{i \mathbf{k} \cdot \mathbf{x}} &= -(2\pi)^{-3} \int d^3 k_1 d^3 k_2 \nabla_i [\delta_{k_1} v_{k_2}^i e^{i(\mathbf{k}_1 + \mathbf{k}_2) \cdot \mathbf{x}}], \\ &= -i(2\pi)^{-3} \int d^3 k_1 d^3 k_2 \delta_{k_1} v_{k_2}^i (k_{1i} + k_{2i}) e^{i(\mathbf{k}_1 + \mathbf{k}_2) \cdot \mathbf{x}} \\ &= -\mathcal{H}(2\pi)^{-3} \int d^3 k_1 d^3 k_2 \delta_{k_1} \theta_{k_2} \frac{k_2^i}{k_2^2} (k_{1i} + k_{2i}) e^{i(\mathbf{k}_1 + \mathbf{k}_2) \cdot \mathbf{x}} \\ &= -\mathcal{H} A(\mathbf{k}_{12}), \end{aligned} \quad (12.31)$$

where $\mathbf{k}_{12} = \mathbf{k}_1 + \mathbf{k}_2$ and

$$A(\mathbf{k}_{12}) = (2\pi)^{-3} \int d^3 k_1 d^3 k_2 \delta_{k_1} \theta_{k_2} \frac{\mathbf{k}_{12} \cdot \mathbf{k}_2}{k_2^2} e^{i \mathbf{k}_{12} \cdot \mathbf{x}}. \quad (12.32)$$

We can introduce the Dirac delta function δ_D to write

$$A(\mathbf{k}_{12}) = \int d^3 k A(\mathbf{k}) \delta_D(\mathbf{k} - \mathbf{k}_{12}). \quad (12.33)$$

Therefore we have

$$\int d^3 k e^{i \mathbf{k} \cdot \mathbf{x}} \mathcal{H} \left[\delta'_k + \theta_k + \int d^3 k_1 d^3 k_2 \delta_{k_1} \theta_{k_2} \frac{\mathbf{k} \cdot \mathbf{k}_2}{k_2^2} \delta_D(\mathbf{k} - \mathbf{k}_{12}) \right] = 0. \quad (12.34)$$

The term inside square brackets has to vanish on account of the completeness and unicity of the Fourier expansion (of course the same property has been silently

¹ The volume factor V in the Fourier pre-factor can be conveniently put to unity when performing a series of Fourier transformations.

used all along the treatment of linear perturbations). We then arrive at the Fourier space version of the continuity equation to second-order

$$\delta'_k + \theta_k = - \int d^3k_1 d^3k_2 \delta_{k_1} \theta_{k_2} \frac{\mathbf{k}}{2} \cdot \left(\frac{\mathbf{k}_1}{k_1^2} + \frac{\mathbf{k}_2}{k_2^2} \right) \delta_D(\mathbf{k} - \mathbf{k}_{12}). \quad (12.35)$$

Similar steps lead directly to the Fourier-transformed Euler equation:

$$\theta'_k + \left(1 + \frac{\mathcal{H}'}{\mathcal{H}} \right) \theta_k + \frac{3}{2} \Omega_m \delta_k = - \int d^3k_1 d^3k_2 \theta_{k_1} \theta_{k_2} \frac{k_{12}^2 (\mathbf{k}_1 \cdot \mathbf{k}_2)}{2k_1^2 k_2^2} \delta_D(\mathbf{k} - \mathbf{k}_{12}). \quad (12.36)$$

Since there is symmetry between $\mathbf{k}_1$ and $\mathbf{k}_2$, we used the substitutions

$$\frac{\mathbf{k} \cdot \mathbf{k}_2}{k_2^2} \rightarrow \frac{\mathbf{k}}{2} \cdot \left(\frac{\mathbf{k}_1}{k_1^2} + \frac{\mathbf{k}_2}{k_2^2} \right), \quad (12.37)$$

and

$$\begin{aligned} (\mathbf{k}_1 \cdot \mathbf{k}_2) \mathbf{k}_2 \cdot (\mathbf{k}_1 + \mathbf{k}_2) &\rightarrow \frac{1}{2} [(\mathbf{k}_1 \cdot \mathbf{k}_2) \mathbf{k}_2 \cdot (\mathbf{k}_1 + \mathbf{k}_2) + (\mathbf{k}_1 \cdot \mathbf{k}_2) \mathbf{k}_1 \cdot (\mathbf{k}_1 + \mathbf{k}_2)] \\ &= \frac{1}{2} k_{12}^2 (\mathbf{k}_1 \cdot \mathbf{k}_2), \end{aligned} \quad (12.38)$$

to symmetrize the result. The solution of these coupled equations can be obtained numerically [500, 501], but for an Einstein–de Sitter Universe we can proceed in the following way. Since we know already from the previous section that the n -th order growth function D_n grows like D_L^n , we can expand the general growing solution in the form

$$\delta(\mathbf{k}, a) = \sum_n D_L^n \delta_n(\mathbf{k}), \quad \theta(\mathbf{k}, a) = - \sum_n D_L^n \theta_n(\mathbf{k}). \quad (12.39)$$

Then we see that at any order the time-dependent factors D_L^n in Eqs. (12.35, 12.36) can be factored out and we can deal with the geometric (k -dependent) factors alone. We can now differentiate δ'_k in Eq. (12.35) and replace θ_k, θ'_k by making use of Eq. (12.36). Then we use the first-order solutions

$$\delta_{k_1} = D_L \delta_1(\mathbf{k}_1), \quad \theta_{k_1} = -D'_L \delta_1(\mathbf{k}_1) = -D_L \delta_1(\mathbf{k}_1), \quad (12.40)$$

(remember that $D_L = D'_L = a$) and similarly for $\mathbf{k}_2$ to obtain, for $n = 2$,

$$\begin{aligned} &\left[(D_L^2)'' + \frac{1}{2} (D_L^2)' - \frac{3}{2} D_L^2 \right] \delta_2(\mathbf{k}) \\ &= D_L^2 \int d^3k_1 d^3k_2 \delta_D(\mathbf{k} - \mathbf{k}_{12}) \delta_1(\mathbf{k}_1) \delta_1(\mathbf{k}_2) \left[\frac{5}{4} \mathbf{k} \cdot \left(\frac{\mathbf{k}_1}{k_1^2} + \frac{\mathbf{k}_2}{k_2^2} \right) + \frac{k_{12}^2 (\mathbf{k}_1 \cdot \mathbf{k}_2)}{2k_1^2 k_2^2} \right], \end{aligned} \quad (12.41)$$

where we used $\Omega_m = 1$ and $\mathcal{H}'/\mathcal{H} = -1/2$. Solving for δ_2 and repeating for θ_2 we obtain the following result [502]

$$\delta_2(\mathbf{k}) = \int d^3k_1 d^3k_2 \delta_D(\mathbf{k} - \mathbf{k}_1 - \mathbf{k}_2) F_2(\mathbf{k}_1, \mathbf{k}_2) \delta_1(\mathbf{k}_1) \delta_1(\mathbf{k}_2), \quad (12.42)$$

$$\theta_2(\mathbf{k}) = \int d^3k_1 d^3k_2 \delta_D(\mathbf{k} - \mathbf{k}_1 - \mathbf{k}_2) G_2(\mathbf{k}_1, \mathbf{k}_2) \theta_1(\mathbf{k}_1) \theta_1(\mathbf{k}_2), \quad (12.43)$$

where

$$F_2 = \frac{5}{7} + \frac{1}{2} \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1 k_2} \left(\frac{k_1}{k_2} + \frac{k_2}{k_1} \right) + \frac{2}{7} \frac{(\mathbf{k}_1 \cdot \mathbf{k}_2)^2}{k_1^2 k_2^2}, \quad (12.44)$$

$$G_2 = \frac{3}{7} + \frac{1}{2} \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1 k_2} \left(\frac{k_1}{k_2} + \frac{k_2}{k_1} \right) + \frac{4}{7} \frac{(\mathbf{k}_1 \cdot \mathbf{k}_2)^2}{k_1^2 k_2^2}. \quad (12.45)$$

By the same procedure, the kernels F_n, G_n can be calculated recursively at all orders and generalized to non-flat spaces (see e.g., Refs. [503, 504] and the review [505]).

A direct way to compare the theoretical result with observations is to estimate the third-order moment or skewness, $\langle \delta^3 \rangle$. From Eq. (12.39) we have

$$\langle \delta^3(\mathbf{k}, a) \rangle = \langle (D_L \delta_1 + D_L^2 \delta_2 + \dots)^3 \rangle \simeq \langle \delta_1^3 \rangle D_L^3 + 3 \langle \delta_1^2 \delta_2 \rangle D_L^4 + \dots \quad (12.46)$$

The first term vanishes because the initial field is Gaussian and its odd moments are zero. Then we have

$$\begin{aligned} \langle \delta^3(\mathbf{k}, a) \rangle &\simeq 3 \langle \delta_1^2 \delta_2 \rangle D_L^4 \\ &= D_L^4 \int d^3k_1 d^3k_2 d^3k_3 d^3k_4 \delta_D(\mathbf{k} - \mathbf{k}_1 - \mathbf{k}_2) \\ &\quad \times F_2(\mathbf{k}_1, \mathbf{k}_2) \langle \delta_1(\mathbf{k}_1) \delta_1(\mathbf{k}_2) \delta_1(\mathbf{k}_3) \delta_1(\mathbf{k}_4) \rangle. \end{aligned}$$

Notice that the first non-trivial moment is then of fourth order in D_L , rather than third as one could have naively expected. The correlation term $\langle \dots \rangle$ for Gaussian variables vanishes except for identical $\mathbf{k}$ pairs, since in a Gaussian field different modes are independent. So we can have non-zero terms only when $\mathbf{k}_1 = \mathbf{k}_2$ or when

$$\langle \delta_1(\mathbf{k}_1) \delta_1(\mathbf{k}_1) \delta_1(\mathbf{k}_3) \delta_1(\mathbf{k}_3) \rangle = \langle \delta_1(\mathbf{k}_1) \delta_1(\mathbf{k}_1) \rangle \langle \delta_1(\mathbf{k}_3) \delta_1(\mathbf{k}_3) \rangle. \quad (12.47)$$

Since $\langle \delta_2(\mathbf{k}) \rangle = 0$, however, the integral over the pair $\mathbf{k}_3 = \mathbf{k}_4$ must vanish. Moreover, the two other possible pairs are identical:

$$\begin{aligned} \langle \delta_1(\mathbf{k}_1) \delta_1(\mathbf{k}_2) \delta_1(\mathbf{k}_1) \delta_1(\mathbf{k}_2) \rangle &= \langle \delta_1(\mathbf{k}_1) \delta_1(\mathbf{k}_2) \delta_1(\mathbf{k}_2) \delta_1(\mathbf{k}_1) \rangle \\ &= \langle \delta_1(\mathbf{k}_1) \delta_1(\mathbf{k}_1) \rangle \langle \delta_1(\mathbf{k}_2) \delta_1(\mathbf{k}_2) \rangle. \end{aligned} \quad (12.48)$$

This shows that $\langle \delta_1(\mathbf{k}_1)\delta_1(\mathbf{k}_2)\delta_1(\mathbf{k}_3)\delta_1(\mathbf{k}_4) \rangle = 2\delta_D(\mathbf{k}_1 - \mathbf{k}_3)\delta_D(\mathbf{k}_2 - \mathbf{k}_4)P(\mathbf{k}_1)P(\mathbf{k}_2)$, where P is the power spectrum. Then we obtain (here we assume an isotropic spectrum)

$$\begin{aligned} \langle \delta^3(\mathbf{k}, a) \rangle &= 3\langle \delta_1^2 \delta_2 \rangle D_L^4 = 6D_L^4 \int d^3k_1 d^3k_2 e^{i(\mathbf{k}_1 - \mathbf{k}_2) \cdot \mathbf{x}} F_2(\mathbf{k}_1, \mathbf{k}_2) P(k_1) P(k_2) \\ &= 6D_L^4 \int k_1^2 dk_1 k_2^2 dk_2 I_{k_1} I_{k_2} P(k_1) P(k_2) \int d\Omega_1 d\Omega_2 F_2(\mathbf{k}_1, \mathbf{k}_2), \end{aligned} \quad (12.49)$$

where $I_k = \sin(kx)/kx$. Similarly, we have

$$\langle \delta^2(\mathbf{k}, a) \rangle = 4\pi D_L^2 \int k_1^2 dk_1 I_{k_1} P(k_1). \quad (12.50)$$

Since $F_2(\mathbf{k}_1, \mathbf{k}_2)$ only depends on the moduli and on the cosine μ_{12} of the $\mathbf{k}_1, \mathbf{k}_2$ angle, the angular part in Eq. (12.49) can be evaluated analytically. Then the skewness for a matter-dominated flat Universe is given by [95]

$$S_3 \equiv \frac{\langle \delta^3 \rangle}{\langle \delta^2 \rangle^2} = \frac{3}{8\pi^2} \int d\Omega_1 d\Omega_2 F_2(\mu_{12}, k_1, k_2) = 3 \int_{-1}^1 d\mu_{12} F_2(\mu_{12}, k_1, k_2) = \frac{34}{7}, \quad (12.51)$$

where we have used the expression $F_2 = 5/7 + (\mu_{12}/2)(k_1/k_2 + k_2/k_1) + (2/7)\mu_{12}^2$ and the identity

$$\int d\Omega_1 d\Omega_2 f(\mu_{12}, k_1, k_2) = 8\pi^2 \int d\mu_{12} f(\mu_{12}, k_1, k_2). \quad (12.52)$$

Similar expressions can be given to any desired order.

An expression that applies to the general solution Eq. (12.26) is [506]

$$F_2 = \frac{1}{2} \left[1 + \alpha + \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1 k_2} \left(\frac{k_1}{k_2} + \frac{k_2}{k_1} \right) + (1 - \alpha) \frac{(\mathbf{k}_1 \cdot \mathbf{k}_2)^2}{k_1^2 k_2^2} \right], \quad (12.53)$$

where $\alpha = D_{2a}/D_L^2$. For the Einstein–de Sitter model the solution to Eq. (12.24) is given by $D_{2a} = (3/7)D_L^2$ and hence $\alpha = 3/7$. Then the skewness can be evaluated as

$$S_3 = 4 + 2\alpha = \frac{34}{7} + \frac{6}{7} \left(\frac{7}{3}\alpha - 1 \right). \quad (12.54)$$

In Ref. [507] the generalization for a single dark energy component with an equation of state $w_{DE} = \text{constant}$ is given.

The observable S_3 is however different from the theoretical calculation above because the continuous density field δ can be obtained only by smoothing the

galaxy counts over some region of space. In practice, this means that what we observe is not $\delta(r)$ but rather

$$\delta_R^{(1)} = \int d^3k \delta_k e^{i\mathbf{k}\cdot\mathbf{r}} W(kR), \quad (12.55)$$

at first-order and

$$\delta_R^{(2)} = \int d^3k \delta_{k_1} \delta_{k_2} F_2(\mathbf{k}_1, \mathbf{k}_2) e^{i(\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{r}} W(|\mathbf{k}_1 + \mathbf{k}_2|R), \quad (12.56)$$

at second-order. Note that $W(kR)$ is the Fourier transform of the window function with a characteristic scale R (see Section 3.3). One may think that the overall effect of the smoothing will amount to a minor correction to S_3 , but this is not the case. The reason is that smoothing a non-Gaussian field is likely to *reduce* the overall non-Gaussianity since it is effectively a way to average over many random variables and by the central limit theorem we can expect the average to be more Gaussian than the individual components. Strictly speaking the limit theorem applies only to independent variables, while here the field $\delta(x)$ is supposed to be correlated. Still the overall effect of the smoothing is indeed to reduce the skewness, at least for not too steep spectra. The remarkably simple result is [500]

$$S_3^{(R)} = S_3 + \frac{d \ln \sigma^2(R)}{d \ln R}, \quad (12.57)$$

where $\sigma^2(R)$ is the variance in cells defined in Eq. (3.58). For a power-law spectrum $P \propto k^n$ the correction is simply given by $-(n+3)$ [508]. For instance, on the scales $R \approx 10 h^{-1} \text{Mpc}$, one can approximate $P \propto k^{-1.5}$ and the correction reduces S_3 by as much as one-third.

Equation (12.24) can be solved for any dark energy model. In uncoupled dark energy models all one has to do is to input the appropriate functions $\Omega_m(a)$, $\mathcal{H}(a)$. It turns out that S_3 has a very weak dependence on $\Omega_m^{(0)}$. In Ref. [505] the following expression is reported

$$S_3 = \frac{34}{7} + \frac{6}{7} [(\Omega_m^{(0)})^{-0.03} - 1]. \quad (12.58)$$

In Ref. [509] a very weak dependence on the dark energy equation of state is also reported, no more than a few percent for realistic w_{DE} 's.

Equation (12.27) can also be generalized to coupled dark energy models. If one compares Eqs. (12.11)–(12.13) with their corresponding equations in Section 11.3, one realizes that the interacting model introduces two simple modifications: a modified friction term in the Euler equation [see Eq. (11.92)] and a modified Poisson equation [see Eq. (11.95)]. These reflect immediately onto the non-linear

equation (12.27) which becomes

$$D_2'' + F(a)D_2' - \frac{3}{2}\Omega_m S(a)D_2 = \frac{4}{3}D_L'^2 + \frac{3}{2}\Omega_m S(a)D_L^2, \quad (12.59)$$

where

$$F(a) = 1 + \frac{\mathcal{H}'}{\mathcal{H}} + Q\phi', \quad S(a) = 1 + 2Q^2, \quad (12.60)$$

to be coupled to the background equations that give ϕ , $\mathcal{H}$, Ω_m as functions of time. In Ref. [487] it has been found that the skewness depends on Q in a weak but non-negligible way as long as the couplings between dark energy and non-relativistic matter (dark matter, baryons) are not the same. For $f(R)$ gravity and the scalar-tensor gravity models we have discussed in Chapter 9, it was shown instead in Ref. [510] that the skewness is not much different from the value in the Λ CDM model.

We can now use the expressions above to evaluate the higher-order corrections to the power spectrum. Inserting Eq. (12.39) into the spectrum definition (pay attention that the prime in $\mathbf{k}'$ does not denote differentiation and that we are putting $V = 1$)

$$\langle \delta(\mathbf{k}, a)\delta(\mathbf{k}', a) \rangle = \delta_D(\mathbf{k} - \mathbf{k}')P(\mathbf{k}, a), \quad (12.61)$$

we obtain to the first non-trivial order (i.e. fourth order in δ)

$$\begin{aligned} \delta_D(\mathbf{k} - \mathbf{k}')P(\mathbf{k}, a) &= D_L^2 \langle \delta_1(\mathbf{k})\delta_1(\mathbf{k}') \rangle \\ &+ D_L^4 [\langle \delta_3(\mathbf{k})\delta_1(\mathbf{k}') \rangle + \langle \delta_1(\mathbf{k})\delta_3(\mathbf{k}') \rangle + \langle \delta_2(\mathbf{k})\delta_2(\mathbf{k}') \rangle]. \end{aligned} \quad (12.62)$$

Defining

$$D_L^{n+m} \langle \delta_n(\mathbf{k}, a)\delta_m(\mathbf{k}', a) \rangle = \delta_D(\mathbf{k} - \mathbf{k}')P_{nm}(\mathbf{k}, a), \quad (12.63)$$

we can write the correction as

$$P(\mathbf{k}, a) = P_L + 2P_{13} + P_{22}, \quad (12.64)$$

where $P_L = P_{11}$ is the “linear” spectrum. Inserting δ_2 from Eq. (12.42) and δ_3 (that we have not written out explicitly) into P_{22} and P_{13} , we find (see e.g., [505])

$$P_{22}(\mathbf{k}, a) = 2 \int [F_2^s(\mathbf{k} - \mathbf{k}', \mathbf{k}')]^2 P_L(|\mathbf{k} - \mathbf{k}'|, a) P_L(\mathbf{k}', a) d^3k', \quad (12.65)$$

$$P_{13}(\mathbf{k}, a) = 3P_L(\mathbf{k}, a) \int F_3^s(\mathbf{k}, \mathbf{k}', -\mathbf{k}') P_L(\mathbf{k}', a) d^3k', \quad (12.66)$$

where the subscript s indicates the symmetrized versions of the kernels.

Further progress may be achieved only numerically and we refer the interested reader to the literature, for instance the review [505]. Naturally, the endeavor of

estimating non-linear corrections to the power spectrum is of the highest importance, given that the $P(k)$ contains so much information on cosmology, as we have already seen and will see more in Chapter 14. However, it must be said that the comparison of the higher-order corrections to N -body simulations of the present Universe is not totally satisfactory (see e.g., Ref. [501]). The fact that higher-order terms in $P(k)$ make a significant contribution signals that the perturbative treatment is about to fail. So the practical usefulness of analytical higher-order corrections to $P(k)$ appears limited; new methods based on renormalization groups have been shown to improve upon standard perturbation theory, see Refs. [511, 512, 513]. An alternative way, to which nowadays most research is devoted, is to find empirical fits of N -body simulations; the relevant literature is cited in Section 12.5.

The concept of the power spectrum may be extended to higher-order correlations. So in place of Eq. (12.61) one writes

$$\langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3) \rangle = \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3)B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3), \quad (12.67)$$

which defines the *bispectrum* B . Since the odd powers of δ_1 vanish on account of Gaussianity, the leading term is [using Eq. (12.39)]

$$\langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3) \rangle = D_L^4 \langle \delta_1(\mathbf{k}_1)\delta_1(\mathbf{k}_2)\delta_2(\mathbf{k}_3) \rangle + \text{cyclic}(231, 321), \quad (12.68)$$

from which we derive

$$B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = 2P_L(\mathbf{k}_1)P_L(\mathbf{k}_2)F_s^{(s)}(\mathbf{k}_1, \mathbf{k}_2) + \text{cyclic}(23, 31). \quad (12.69)$$

The bispectrum finds an interesting application as a tool to estimate the bias between the observed galaxy δ and the underlying δ_m , if we assume a scale-independent bias, since it can be shown that the bispectrum depends on the bias in a way that varies with the shape of the triangles $\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3 = -\mathbf{k}_1 - \mathbf{k}_2$ in Fourier space [514]. Estimating the bispectrum for several shapes allows then to constrain the bias. This method has been discussed and applied to real data in Ref. [515]. The potential of the bispectrum for estimating dark energy parameters has still to be investigated in detail.

12.3 Spherical collapse

The full equation (12.20) corresponds to the evolution of the density contrast in a spherical perturbation. It can be derived on purely Newtonian grounds by noting that a shell of matter at distance R from the center of a spherical overdensity with uniform density ρ moves according to the Newtonian force law

$$\frac{d^2 R}{dt^2} = -\frac{GM(R)}{R^2} = -\frac{4}{3}\pi G\rho R, \quad (12.70)$$

where $M(R) = 4\pi\rho R^3/3$ is the *constant* mass inside the shell. Since for pressureless matter the background density scales as $\rho_0 = (3M(R_0)/4\pi)(R_0 a(t))^{-3}$, where R_0 is the initial size of the perturbation, we can define the density contrast as

$$\delta = \left(\frac{a(t)R_0}{R} \right)^3 - 1, \quad (12.71)$$

inside the shell and $\delta = 0$ outside. The crucial assumption here is that δ is a step, or *top-hat*, function, which in fact allows all spatial derivatives to be cancelled. The equation for δ in our time variable N is then exactly as in Eq. (12.20):

$$\delta'' + \left(1 + \frac{\mathcal{H}'}{\mathcal{H}} \right) \delta' - \frac{3}{2} \Omega_m \delta = \frac{4}{3} \frac{\delta'^2}{1 + \delta} + \frac{3}{2} \Omega_m \delta^2. \quad (12.72)$$

Of course, this is nothing more than a consequence of the fact that Eq. (12.20) has been derived for a shear-free fluid and therefore an initially spherical perturbation remains spherical [516].

Multiplying Eq. (12.70) on both sides by $2dR/dt$ the equation can be integrated once as

$$\left(\frac{dR}{dt} \right)^2 = \frac{2GM}{R} - C, \quad (12.73)$$

where C is an integration constant. This is the cycloid equation, whose solution for $C > 0$ can be given parametrically as $R = GM(1 - \cos \tau)/C$ and $t = GM(\tau - \sin \tau)/C^{3/2}$ where $\tau \in (0, 2\pi)$. Substituting in δ and putting $a(t) = a_0(t/t_0)^{2/3}$ we obtain in the Einstein–de Sitter case:

$$\delta = \frac{9}{2} \frac{(\tau - \sin \tau)^2}{(1 - \cos \tau)^3} - 1, \quad (12.74)$$

$$\delta_L = \frac{3}{5} \left[\frac{3}{4} (\tau - \sin \tau) \right]^{2/3}, \quad (12.75)$$

where $\delta_L (> 0)$ is the solution of the linearized equation and where the integration constant C has been chosen to set $\delta(\tau = 0) = 0$. It is convenient to use δ_L as a bookkeeping device: we express the behavior of δ as a function of δ_L instead of the parameter τ . A similar solution exists for an underdensity $\delta_L < 0$. We have assumed a constant mass $M(R)$: this implies that our analysis is valid only until shell-crossing occurs. As one expects, the radius R first increases (a small perturbation expands with the cosmological expansion), reaches a turnaround point and then decreases to zero (the perturbation collapses under its own gravity). The final singular phase is of course unphysical because the dust assumption will fail at some high density, non-radial fluctuations will develop and even the dark matter

collisionless component will undergo the so-called “violent relaxation” mechanism and will set into virial equilibrium.

The main result we get from this model is the *critical* or *collapse* value δ_{coll} of the *linear* fluctuation δ_L that is reached at the time of collapse. This quantity is of cosmological relevance because it is used in the Press–Schechter theory [517, 518] as a first approximation to the epoch of galaxy formation and to calculate the abundance of collapsed objects, as we will discuss below. It can be seen from Eq. (12.75) that when $\tau = 0$ the perturbations are zero, then δ reaches a turnaround at $\tau = \pi$ (for which $\delta_T \equiv \delta(\pi) = (3\pi/4)^2 - 1 \approx 4.6$ and $\delta_L \approx 1.063$) and finally for $\tau = 2\pi$ the overdensity δ (but of course not δ_L) becomes singular. This singularity occurs when

$$\delta_L = \delta_{\text{coll}} = (3/5)(3\pi/2)^{2/3} \approx 1.686, \quad (12.76)$$

and it takes exactly twice as much time as for the turnaround. Notice that this value is independent of time: a spherical perturbation in the Einstein–de Sitter Universe collapses to a singularity whenever the linear density contrast equals 1.686. For other models, however, δ_{coll} depends on time. An approximation for dark energy with constant w_{DE} in flat space is [519]

$$\delta_{\text{coll}}(z) = 1.686 \left[1 + \alpha(w_{\text{DE}}) \log_{10} \Omega_m(z) \right], \quad (12.77)$$

$$\alpha(w_{\text{DE}}) = 0.353w_{\text{DE}}^4 + 1.044w_{\text{DE}}^3 + 1.128w_{\text{DE}}^2 + 0.555w_{\text{DE}} + 0.131. \quad (12.78)$$

One can define other phenomenologically interesting epochs that are sometimes used: the epoch of non-linearity ($\delta = 1$, corresponding to $\delta_L \approx 0.57$) and the epoch of *expected* virialization. The latter is *defined* to correspond to the instant in which the kinetic energy K is related to the gravitational potential energy U by the condition

$$K = \frac{R}{2} \frac{\partial U}{\partial R}. \quad (12.79)$$

However, it is by no means obvious that this condition is enough to realize virialization, especially when dark energy is present. For an inverse-power potential ($U \propto -1/R$), the virialization implies $K = -U/2$. The radius and the density of the perturbation at virialization can be calculated by assuming conservation of energy at turnaround (when the kinetic energy vanishes; subscript T) and at a virialization epoch t_V when the kinetic energy satisfies $K_V = -U_V/2$, i.e.

$$U_T = U_V + K_V = U_V/2. \quad (12.80)$$

Since for a uniform sphere $U = -3GM/5R$ (and remembering once again we are assuming $M = \text{constant}$), we obtain the relation $R_V = R_T/2$. Hence the virialized radius is half the turnaround radius. The density inside this radius turns

out to be $\delta_V \approx 178$ and the epoch of this occurrence is very close to the final collapse time (see problem 12.3). A numerical fit for $w_{\text{DE}} = \text{constant}$ models in flat space gives [519]

$$\delta_V \approx 178[1 + b_1 \theta^{b_2}(z)], \quad (12.81)$$

$$\theta \equiv \frac{1 - \Omega_m(z)}{\Omega_m(z)}, \quad (12.82)$$

$$b_1 \equiv 0.399 - 1.309(|w_{\text{DE}}|^{0.426} - 1), \quad (12.83)$$

$$b_2 \equiv 0.941 - 0.205(|w_{\text{DE}}|^{0.938} - 1), \quad (12.84)$$

if z is the collapse redshift.

It is not easy to extend this method to a general dark energy model because the clustering properties of dark energy are quite difficult to deal with, even at the linear order. For instance, even if dark energy clusters on the relevant scales, it is not obvious whether it also virializes, whether it virializes at the same epoch as matter, and whether we can assume energy conservation. If we simply assume dark energy does not cluster at all then one can simply generalize Eq. (12.70) to

$$\frac{d^2 R}{dt^2} = -\frac{GM(R)}{R^2} = -\frac{4}{3}\pi G [\rho_m + (1 + w_{\text{DE}})\rho_{\text{DE}}] R, \quad (12.85)$$

where for a constant equation of state w_{DE} we have $\rho_{\text{DE}} \sim a^{-3(1+w_{\text{DE}})}$ both inside and outside the overdensity and as before $\rho_m \propto R^{-3}$. Then we obtain Eq. (12.72) again, where $(d\mathcal{H}/dN)/\mathcal{H} = -(1 + 3w_{\text{eff}})/2 \simeq -(1 + 3\Omega_{\text{DE}}w_{\text{DE}})/2$. The critical value δ_{coll} must now be obtained numerically. In this case the correct initial condition is found by assuming zero initial velocity for the shells. This corresponds to a mixture of growing and decaying modes $\delta_{\pm}$, so that the initial perturbation corresponds to $\delta_i = (5/3)\delta_+$. A range of values $\delta_{\text{coll}} = 1.6\text{--}1.686$ was found in Ref. [478], with little dependence for constant w_{DE} in the range $(-1, -1/3)$.

If dark energy does not cluster we have $\rho_{\text{DE}} \sim a^{-3(1+w_{\text{DE}})}$; if it clusters just as a perfect fluid with a constant equation of state w_{DE} then we can expect $\rho_{\text{DE}} \sim R^{-3(1+w_{\text{DE}})}$. Some authors suggested that a tentative way to take into account dark energy clustering is to parametrize the transition between the two forms as [520, 521]

$$\rho_{\text{DE}} \sim R^{-3(1-\beta)(1+w_{\text{DE}})} a^{-3\beta(1+w_{\text{DE}})}, \quad (12.86)$$

so that $\beta = 1$ means dark energy does not cluster at all, $\beta = 0$ means matter-like clustering, and intermediate β means some intermediate form of clustering. The main point of this parametrization is to demonstrate that the values of R_V and δ_V depend sensitively on the dark energy properties. When matter alone virializes, the final R_V/R_T ratio is smaller than $1/2$. When both matter and dark energy virialize, the final ratio is larger. Other authors generalized Eq. (12.72) to fluids with finite

sound speed, assuming that a top-hat perturbation remains a good approximation [522]. Spherical collapse in coupled dark energy models has been studied in Ref. [523].

It is difficult to go much beyond this kind of phenomenological parametrization. A full understanding of non-linear physics in dark energy would require extensive N -body simulations coupled to lattice simulations of scalar fields, a technical feat which is still largely to be explored.

12.4 The mass function of collapsed objects

The main reason why it is worthwhile to discuss the abstract phenomenon of a “spherical collapse” is that the critical value δ_{coll} and the virial radius R_V (or rather the mass contained within that radius) enter the Press–Schechter (PS) formula for the abundance of virialized objects. The main idea behind the PS formula is that we can estimate the number of collapsed objects formed in a random Gaussian field by simply counting at any given time how many regions have an overdensity above the collapse threshold given by δ_{coll} .

Suppose at some redshift z we smooth a random Gaussian field of density fluctuations over cells of radius R , each containing on average the mass $M = 4\pi R^3 \rho/3$ with $\rho(z)$ the background density. Since the smoothing is a linear operation, if the field is Gaussian then also the density contrast δ in the cells will be distributed as a Gaussian probability distribution function with variance $\sigma_M^2(z)$. Suppose that *all* the cells with $\delta > \delta_{\text{coll}}$ undergo collapse and virialization. The fraction of collapsed regions (i.e. the fraction of space containing objects of mass *larger* than M) will be then

$$p(M, z)|_{\delta > \delta_{\text{coll}}} = \frac{1}{\sigma_M(z)\sqrt{2\pi}} \int_{\delta_{\text{coll}}}^{\infty} \exp\left(-\frac{\delta_M^2}{2\sigma_M^2(z)}\right) d\delta_M = \frac{1}{2} \text{erfc}\left(\frac{\delta_{\text{coll}}}{\sqrt{2}\sigma_M(z)}\right), \quad (12.87)$$

where $\text{erfc}(x)$ is the error function. The fraction containing objects of mass within the range $[M, M + dM]$ is given by

$$dp(M, z) = \left| \frac{\partial p(M, z)|_{\delta > \delta_{\text{coll}}}}{\partial M} \right| dM. \quad (12.88)$$

Remember that in general the threshold δ_{coll} depends on z . Although the boxes with $\delta > \delta_{\text{coll}}$ are certainly not in the linear regime, the idea is to use the linear regime to estimate the fraction of collapsed regions. We are then implicitly assuming that the variance $\sigma_M(z)$ is in the linear regime ($\sigma_M \ll 1$) and therefore that it can be calculated from Eq. (3.58) with the linear spectrum at any redshift. By using the growth function $D(z)$ we have $\sigma_M(z) = D(z)\sigma_M(0)$.

Now, suppose in a volume V we find N collapsed objects, each occupying a volume $V_M = M/\rho$. Then by definition the volume occupied collectively by the N objects is the fraction dp of V , i.e.

$$NV_M = Vdp, \quad (12.89)$$

and therefore the number density dn of collapsed halos with mass in the dM range (the *mass function*) will be

$$\begin{aligned} dn &= \frac{N}{V} = \frac{dp}{V_M} = \frac{\rho}{M} \left| \frac{\partial p(M, z)|_{\delta > \delta_{\text{coll}}}}{\partial M} \right| dM \\ &= \sqrt{\frac{2}{\pi}} \frac{\rho}{M^2} \frac{\delta_{\text{coll}}}{\sigma_M} \left| \frac{d \ln \sigma_M}{d \ln M} \right| e^{-\delta_{\text{coll}}^2/(2\sigma_M^2)} dM. \end{aligned} \quad (12.90)$$

The extra factor of two that we have inserted in the last step is required because we want all the masses to end up in some object, so that we impose the condition

$$V \int_0^\infty \left(\frac{dn}{dM} \right) dM = 1. \quad (12.91)$$

This factor-of-2 adjustment can be justified with a random walk analysis of fluctuations [524]. In any case, one finds it necessary to fit N -body simulations. Sometimes the number density $n(M, z)$ is taken to be the comoving number density (i.e. is multiplied by a^3): in this case also ρ should be identified with the comoving background density.

Equivalently, Eq. (12.90) is sometimes written as

$$\frac{M}{\rho} \left| \frac{dn}{d \ln \sigma_M} \right| = f(\sigma_M, z), \quad (12.92)$$

where all the cosmological information is contained in the function

$$f(\sigma_M, z) = \sqrt{\frac{2}{\pi}} \frac{\delta_{\text{coll}}}{\sigma_M} e^{-\delta_{\text{coll}}^2/(2\sigma_M^2)}. \quad (12.93)$$

The number density $dn(M, z)$ can then be “directly” confronted with the observed densities of objects (clusters, galaxies, quasars) at any redshift. The most relevant example of this comparison will be discussed in Section 14.5. The mass M is often taken to be the virial mass of that class of objects. Because of the exponential dependence on $\delta_{\text{coll}}/\sigma_M$, the PS formula is quite sensitive to the cosmological model (see Fig. 12.1).

The simplicity of the PS approach must not hide the fact that it relies on a dangerous extrapolation of the linear theory, on the critical assumption of spherical collapse with top-hat filter, on a dubious definition of virialization, and on the

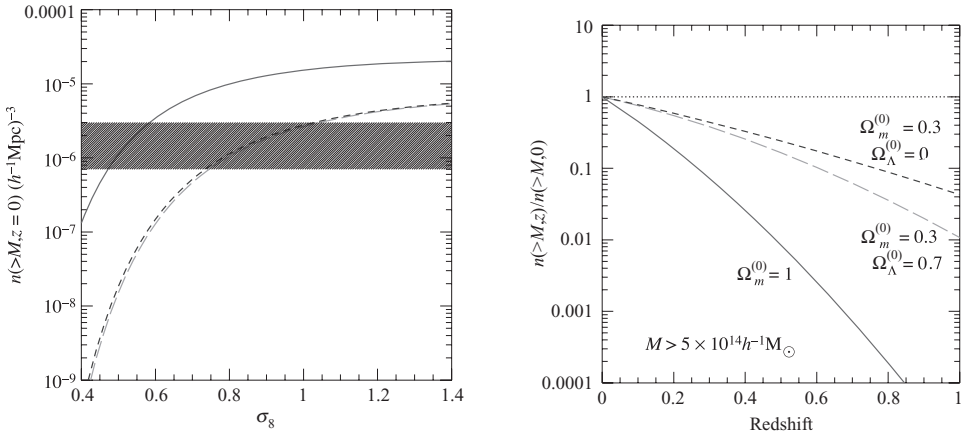


Figure 12.1 The sensitivity of the cluster mass function to cosmological models. Left panel: The cumulative mass function at $z = 0$ for $M > 5 \times 10^{14} h^{-1} M_\odot$ ($M_\odot$ is the solar mass) for three cosmologies, as a function of σ_8 ; solid line: $\Omega_m^{(0)} = 1$; short-dashed line: $\Omega_m^{(0)} = 0.3, \Omega_\Lambda^{(0)} = 0.7$; long-dashed line: $\Omega_m^{(0)} = 0.3, \Omega_\Lambda^{(0)} = 0$. The shaded area indicates the observational uncertainty in the determination of the local cluster space density. Right panel: Evolution of $n(>M, z)$ for the same cosmologies and the same mass limit, with $\sigma_8 = 0.5$ for the $\Omega_m^{(0)} = 1$ case and $\sigma_8 = 0.8$ for the low-density models. From Ref. [528].

absence of processes like merging, dissipation, shell crossing. Surprisingly, this shaky foundation did not prevent the PS formula proving itself a valuable first approximation to the abundances obtained through numerical simulations. Not surprisingly, many works have been dedicated to improving the original PS formula by including corrections due to departure from sphericity [525] or merging [526] or by directly fitting to large N -body simulations. A remarkably successful fit is given by [527]

$$f(\sigma_M, z) = 0.315 \exp(-|0.61 - \ln \sigma_M(z)|^{3.8}). \quad (12.94)$$

This fit has been found to hold for a large range of masses, redshifts, and cosmological parameters, including dark energy with constant or varying w_{DE} [529].

12.5 Dark energy N -body simulations

Beyond the few and uncertain non-linear islands we have seen in the previous sections, lies the frightening ocean of numerical simulations. In that vast ocean, a full treatment of dark energy dynamics is still beyond the horizon, so this section is no more than a sketchy guide to the first explorations.

The idea behind most work on cosmological N -body simulations is to find a general formula that takes as input the linear power spectrum $P_L(k)$ and gives as

output an approximate non-linear version that can be directly compared to data down to relatively small scales:

$$P_{NL}(k) = f_{NL}[P_L(k)]. \quad (12.95)$$

The transformation f_{NL} will in general depend on all the cosmological parameters and is obtained by an empirical fit to the results of the numerical simulations.

Since we have very little clues on the behavior of f_{NL} , except perhaps that we expect the non-linear power spectrum to have more power than the corresponding linear one at galactic scales, most researchers have adopted one of the following two empirical/intuitive approaches. We refer to the cited literature for the full fitting formulae for the non-linear spectra.

The first one, first proposed in Ref. [530] and extended in Ref. [531], is to assume that non-linearity can be well approximated by a remapping of scales: as a collapse brings an initial radius R_0 into a virialized radius R_V , so an initial wavenumber k_L will be carried into a final wavenumber k_{NL} . If we find the relation $k_{NL} = f_{NL}(k_L)$ we could simply write

$$P_{NL}(k_{NL}) = P_L[f_{NL}(k_L)]. \quad (12.96)$$

The second approach is based on the idea that non-linearity mostly concerns objects that form within large dark matter halos. Then we can approximate the halo distribution with the linear spectrum and describe the sub-halo distribution by some “occupation number,” i.e. the number of structures within the same halo. This will be an empirically determined function of the halo density profile. The linear part will dominate at large scales while the sub-halo one will do at small scales. Assuming the processes are essentially independent, the two contributions will simply add in the power spectrum. This program has been first carried out in Ref. [532].

The first work to include explicitly a cosmological constant in the non-linear fitting formula has been done in Ref. [531] using the $k_L \rightarrow k_{NL}$ mapping. In Ref. [532] the fitting formula including the cosmological constant was obtained by using the halo model, see Fig. 12.2. Extending this work to non-clustering dark energy is demanding but straightforward, since the change comes from the background expansion rate [533, 534]. Improvements for dynamical dark energy upon previous fits have been found in Ref. [535]. Other works considered modifications of the Newtonian force inspired by coupled dark energy [536, 537] or by general models of modified gravity [538, 539, 540, 541, 542], still assuming unclustered dark energy. Since this approach changes the force itself, rather than the background expansion, it is clear that it is possible to find important deviations from standard results in terms of halo profile and mass functions.

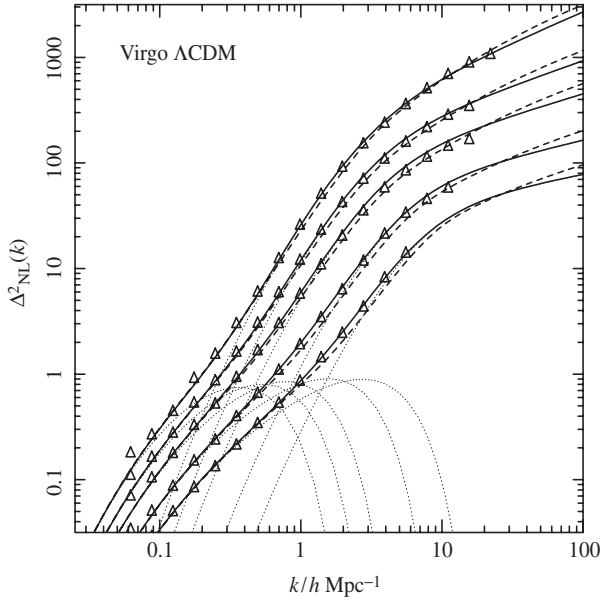


Figure 12.2 The triangles represent the power spectrum $\Delta^2 \sim k^3 P(k)$ of N -body simulations for the Λ CDM model at various epochs $z = 0, 0.5, 1, 2, 3$ (low to high). The full line corresponds to the non-linear fit by Smith *et al.* [532]. The dotted lines are the decomposition into “self-halo” and “halo-halo” terms. The short dashed line corresponds to the old fit of Ref. [531]. From Ref. [532].

The next step is of course to include in full the dark energy clustering. If we model dark energy with a scalar field this will require an N -body simulation coupled to a lattice scalar field simulation, opening a new frontier for high-level computing [539]. This will be more interesting for those models that couple dark energy to matter since here one can expect the matter non-linearity will have a direct and strong influence on dark energy non-linearity. Whether this influence will bear some surprise remains to be seen.

12.6 Problems

- 12.1 Show that Eqs. (12.7)–(12.9) follow from (12.1)–(12.3) under the transformation (12.5).
- 12.2 Derive Eq. (12.22).
- 12.3 Show that the value of the density contrast at virialization in the Einstein–de Sitter Universe is $\delta_V \approx 178$ (see Section 12.3). Assume that the virialization occurs practically at the same time as the final singularity.
- 12.4 What is the conservation equation for ρ_{DE} that reproduces Eq. (12.86)?

13

Statistical methods in cosmology

As a technical introduction to the methods that will be discussed in the next chapter, we review some statistical tools most employed in modern cosmology, e.g., likelihood analysis, Bayes' theorem, model selection, Fisher matrix, and principal component analysis. Note that some of these statistical methods have been implicitly used in previous chapters for observational constraints on dark energy models. The Bayesian approach is particularly suitable for dark energy research because of its flexibility in combining results from different observations and in allowing a direct comparison between various parametrizations.

13.1 The likelihood function

Let us suppose we know, or have good reasons to suspect, that a random variable x , e.g., the apparent magnitude of a supernova, has a probability distribution function (PDF) $f(x; \theta)$ that depends on an *unknown* parameter θ , e.g., the absolute magnitude. The “;” is meant to distinguish the random variables x from the parameter θ . Such a probability is called a *conditional probability* of having the data x given the theoretical parameter θ . We may for instance suppose that the apparent magnitude m is distributed as a Gaussian variable with a given variance σ^2 (the observational error on m), but we do not know one of the cosmological parameters that enter the expected value $m_{\text{th}} = 5 \log_{10} d_L(z; \Omega_m^{(0)}, \Omega_\Lambda^{(0)}) + \text{constant}$, where d_L is the luminosity distance.

If we repeat the measure and we obtain $x_1, x_2, x_3 \dots$, then the law of joint probability tells us that the probability of obtaining x_1 in the interval dx_1 around x_1 , x_2 in the interval dx_2 around x_2 and so forth is

$$f(x_i; \theta) d^n x_i \equiv \prod_i f_i(x_i; \theta) dx_i = f_1(x_1; \theta) f_2(x_2; \theta) f_3(x_3; \theta) \dots dx_1 dx_2 dx_3 \dots, \quad (13.1)$$

if the measures are independent of each other. Clearly, for every θ this multivariate PDF will assume a different value. It is logical to *define* the best θ as the value for which $\prod_i f(x_i; \theta)$ is maximal. Indeed, if we generate random variables distributed as $f(x; \theta)$, the most likely outcome for x is the value that maximizes $f(x; \theta)$. Conversely, if we have a particular outcome x , then our best bet is to assume that θ is such as to maximize the occurrence of that x . We used as an example independent data and a single parameter but this is by no means necessary. We define the best θ_i as those parameters that maximize the joint function $f(x_1, x_2, \dots, x_n; \theta_1, \theta_2, \dots, \theta_m)$. Since in general we have many parameters to estimate, we write the function simply $f(x_i; \theta_j)$, meaning all the x_i 's and all the θ_j 's.

The maximum likelihood method of parameter estimation consists therefore in finding the parameters that maximize the *likelihood function* $f(x_i; \theta_j)$ by solving the system

$$\frac{\partial f(x_i; \theta_j)}{\partial \theta_j} = 0, \quad j = 1, \dots, m. \quad (13.2)$$

Let us denote the solutions of these equations as $\hat{\theta}_i$. They are functions of the data x_i and therefore are random variables, just as the data are. So the classical *frequentist* approach would try to determine the distribution of the $\hat{\theta}_j$ s knowing the distribution of the x_i s; if this is possible, one can assign probabilities to $\hat{\theta}_j$'s ranges, for instance determine the interval of $\hat{\theta}_j$ that contains 95% probability that a particular set of data has been drawn from the theoretical distribution. One problem with this approach is that it is often too difficult to derive the $\hat{\theta}_j$'s distribution analytically and very demanding to derive them numerically through simulated datasets. But the main problem is that this approach does not take into account what we already know concerning the theoretical parameters, for instance the result of previous experiments. To handle this information properly we need to switch to the *Bayesian* approach. Instead of looking for the probability $f(x_i; \theta_j)$ of having the data given the model, we estimate the probability $L(\theta_j; x_i)$ of having the model given the data.

This problem is solved by the fundamental theorem of conditional probabilities, the so-called Bayes' theorem¹:

$$P(T; D) = \frac{P(D; T)P(T)}{P(D)}, \quad (13.3)$$

where we denote the known data x_i with D and the unknown theory (that is, the theoretical parameters θ_j) with T . On the r.h.s., $P(D; T)$ is the conditional probability of having the data given the theory; $P(T)$ and $P(D)$ are the probability of having the theory and the data, respectively; finally, on the l.h.s., $P(T; D)$ is the

¹ Reverend Thomas Bayes (1702–1761) studied what in modern terminology is the binomial distribution and introduced the concept of conditional probability. His work was published posthumously in 1763.

conditional probability of having the theory given the data. Bayes' theorem is a consequence of the definition of conditional probability $P(A; B) \equiv P(A, B)/P(B)$ and of the symmetry of the joint probability $P(A, B)$ (the probability of having both A and B) under the exchange of A, B .

It follows that

$$L(\theta_j; x_i) = \frac{f(x_i; \theta_j)p(\theta_j)}{g(x_i)}, \quad (13.4)$$

where $p(\theta_i)$ is called the *prior* probability for the parameters θ_i , while $g(x_i)$ is the PDF of the data x_i . The final function $L(\theta_j; x_i)$ (or simply $L(\theta_j)$ for shortness) is called *posterior* although sometimes it is also loosely called likelihood just as $f(x_i; \theta_j)$ and generally denoted as L . The posterior contains the information we are looking for: the probability distribution of the parameters given that we observed the data x_i and that we have some prior knowledge about the parameters themselves. In fact the whole method in the Bayesian context should be called “the posterior method” rather than the “likelihood” method.

Since $L(\theta_j; x_i)$ is a probability distribution function for θ_j , it has to be normalized to unity:

$$\int L(\theta_j; x_i) d^n \theta_j = 1 = \frac{\int f(x_i; \theta_j)p(\theta_j) d^n \theta_j}{g(x_i)}, \quad (13.5)$$

and therefore

$$\int f(x_i; \theta_j)p(\theta_j) d^n \theta_j = g(x_i). \quad (13.6)$$

As we will see in the next section the integral on the l.h.s. is called *evidence* and the same name is sometimes given also to $g(x_i)$. The function $g(x_i)$ does not depend on the parameters θ_i and therefore it is of no help in estimating the parameters. From the point of view of $L(\theta_j)$ it is just a normalization factor. The prior $p(\theta_j)$ is also often unknown. Normally we do not know the probability distribution of theories, that is, whether the Λ CDM model is more probable, from an absolute point of view, than a modified gravity model or whether $\Omega_\Lambda^{(0)} = 0$ is more probable than $\Omega_\Lambda^{(0)} = 0.7$. However, we often *do know* something which, while not quite absolute in any sense, still represents some form of information independent of the data at hand. Namely, we know the results of previous experiments. If an experiment convincingly excluded, say, $\Omega_m^{(0)} < 0.1$, then we could use this information, putting $p(\Omega_m^{(0)} < 0.1) = 0$. If instead we believe that $h = 0.72 \pm 0.08$, then we could use as $p(h)$ a Gaussian with mean 0.72 and standard deviation 0.08. These are typical *prior distributions*.

Priors can be of many kinds. Beside including other experiments, we could simply exclude unphysical values, e.g., $\Omega_m^{(0)} < 0$ or weigh down some regions of

parameter space that we, perhaps subjectively, consider less likely. What matters is not so much what we decide to include as prior but rather that we make this decision explicit to the reader and to the potential user of our results. Every posterior, sooner or later, will become a prior for us or for somebody else, and it is our responsibility to make it explicit which prior information we adopted, no less to avoid that a future user includes twice the same information. The easiness of including prior information of all kinds is one of the major advantages of the Bayesian approach.

There are two important facts to note about priors. First, priors matter. Clearly the final result depends on the prior, just as our bet on the result of a football match depends on what we know about the teams based on previous games (and on our personal interpretation of those results). One can say that priors quantify our physical intuition. Second, priors are unavoidable. Even if we are not consciously choosing a prior, the way we manage the statistical problem at hand *always* implies some form of prior. No prior on a parameter means in fact $p(\theta) = 1$ in the domain where θ is defined and $p(\theta) = 0$ outside. Even when θ is defined in the whole real range we are still choosing a “flat” prior, $p(\theta) = 1$, over other possible choices. One could push this argument as far as saying that our choice of theory and its parameters θ already constitute a strong prior. So, again, the important issue is to specify exactly what prior is employed.

Once we have $L(\theta_j)$ we need to search the maximum to obtain the maximum likelihood estimators $\hat{\theta}_i$. Because of the priors, this will differ in general from the maximum of $f(x_i; \theta_j)$. Equation (13.2) is then replaced by

$$\frac{\partial L(\theta_i)}{\partial \theta_i} = 0, \quad i = 1, \dots, n. \quad (13.7)$$

If, as is usually the case, we discard the denominator $g(x_i)$ in Eq. (13.4), the posterior L is not normalized and its normalization has to be recalculated. The overall normalization N is the integral over the parameter space:

$$N = \int L(\theta_i) d^n \theta_i, \quad (13.8)$$

where the integral extends to the whole parameter domain. From the normalized likelihood [i.e. $L(\theta_i)/N$ which we keep calling $L(\theta_i)$], we can derive the regions of confidence (or *belief*) for the parameters. These are defined as regions $R(\alpha)$ delimited by constant $L(\theta_i)$ for which

$$\int_{R(\alpha)} L(\theta_i) d^n \theta = \alpha. \quad (13.9)$$

The region $R(\alpha)$ is the region for which the integral above evaluates to $0 < \alpha < 1$ (remember that now L is normalized and therefore its integral over the whole

domain is 1). To find R one evaluates

$$\int \hat{L}(L_i) d^n \theta = \alpha_i, \quad (13.10)$$

where $\hat{L}(L_i) = L$ if $L > L_i$ and 0 elsewhere (i.e. the volume lying within the curve of “height” L_i , smaller than the peak of L). By trial and error (or by interpolating over a grid of L_i) one finds the preferred α_i . The typical choices are $\alpha = 0.683, 0.954, 0.997$ (also denoted as $1, 2, 3\sigma$, respectively, but sometimes other reference values are also employed. The value L_i that corresponds to α_i is the level at which we have to cut L to find the region $R(\alpha_i)$.

Often, we are interested in some subset of parameters and consider the others as “nuisance” which we would gladly get rid of. For instance, if we are analyzing a set of supernovae apparent magnitudes m_i and comparing them to the theoretical predictions $m_{\text{th}} = 5 \log_{10} d_L(z; \Omega_m^{(0)}, \Omega_\Lambda^{(0)}) + C$, we may be interested in $\Omega_m^{(0)}, \Omega_\Lambda^{(0)}$ but not in the constant C . As we have seen in Section 5.2 this constant depends on the K correction and on the standard absolute magnitude M , to which we can add also the constant $\log_{10} H_0^{-1}$. Our general likelihood is therefore a function of $C, \Omega_m^{(0)}, \Omega_\Lambda^{(0)}$ but we can transform it into a function of $\Omega_m^{(0)}, \Omega_\Lambda^{(0)}$ alone by integrating out C :

$$L(\Omega_m^{(0)}, \Omega_\Lambda^{(0)}) \equiv \int L(C, \Omega_m^{(0)}, \Omega_\Lambda^{(0)}) dC, \quad (13.11)$$

where the integration extends over the domain of definition of C , which in absence of better information could as well be from $-\infty$ to $+\infty$ [there should be no confusion by denoting both the “old” and the “new” likelihood by the same symbol in Eq. (13.11)]. This very common procedure is called *marginalization*.

Often one wants to marginalize a multidimensional problem down to a more manageable and plottable two-dimensional likelihood. Also, one could quote final confidence regions by marginalizing in turn to single parameters, e.g.,

$$L(\Omega_\Lambda^{(0)}) = \int_0^\infty L(\Omega_m^{(0)}, \Omega_\Lambda^{(0)}) d\Omega_m^{(0)}. \quad (13.12)$$

For instance, if the maximum likelihood estimator of $\Omega_m^{(0)}$ is 0.3 and

$$\int_R L(\Omega_m^{(0)}) d\Omega_m^{(0)} = 0.683, \quad (13.13)$$

when R is the interval $\Omega_m^{(0)} = [0.1, 0.4]$, we will write as our final result $\Omega_m^{(0)} = 0.3_{-0.2}^{+0.1}$ at 68.3% confidence level (or, less precisely, at 1σ : notice that this will absolutely not imply that at 2σ one should expect -0.1 as lower limit of $\Omega_m^{(0)}$!).

In the common case in which we want to marginalize over a constant offset or over a multiplicative factor one can often obtain an analytical result. Here we work out the first case, leaving the second to problem 13.2.

Taking again the example of supernovae, suppose that we have N standard candle sources at redshifts z_i with apparent magnitudes m_i and that our preferred cosmological model predicts magnitudes $m_{\text{th},i} = M + 5 \log_{10} d_L(z_i; \theta_j) + 25$, where $d_L(z_i; \theta_j)$ is the luminosity distance measured in Megaparsecs. The luminosity distance is proportional to $1/H_0$. We can therefore take this factor out of the logarithm and write $m_{\text{th},i} = \alpha + \mu_i$, where $\mu_i = 5 \log_{10} \hat{d}_L(z_i; \theta_j)$ and $\alpha = M + 25 - 5 \log_{10} H_0$ and $\hat{d}_L$ is $d_L H_0$. We have very little information on α , so we decide to marginalize it over:

$$L(\theta_j) = N \int d\alpha \exp \left[-\frac{1}{2} \sum_i \frac{(m_i - \mu_i - \alpha)^2}{\sigma_i^2} \right], \quad (13.14)$$

where N is an unimportant normalization factor. Then we have

$$\begin{aligned} L(\theta_j) &= N \int d\alpha \exp \left[-\frac{1}{2} \sum_i \frac{(m_i - \mu_i)^2 + \alpha^2 - 2\alpha(m_i - \mu_i)}{\sigma_i^2} \right] \\ &= N \exp(-S_2/2) \int d\alpha \exp(\alpha S_1 - \alpha^2 S_0/2) \\ &= N \exp \left[-\frac{1}{2} \left(S_2 - \frac{S_1^2}{S_0} \right) \right] \int d\alpha \exp \left[-\frac{1}{2} \left(\alpha - \frac{S_1}{S_0} \right)^2 S_0 \right], \end{aligned} \quad (13.15)$$

where $S_0 = \sum (1/\sigma_i^2)$, $S_1 = \sum y_i/\sigma_i^2$, $S_2 = \sum y_i^2/\sigma_i^2$, and $y_i = m_i - \mu_i$. The integration in the region $(-\infty, +\infty)$ gives a constant independent of μ_i and therefore independent of the theoretical parameters that we absorb in N :

$$L(\theta_j) = N \exp \left[-\frac{1}{2} \left(S_2 - \frac{S_1^2}{S_0} \right) \right]. \quad (13.16)$$

This is then the new likelihood marginalized over the nuisance additive parameter α . Notice that the parameters θ_j ended up inside μ_i which are inside S_1, S_2 . A similar analytic integration can get rid of multiplicative parameters. If the analytical integration is impossible, then one has to marginalize numerically.

Sometimes one prefers to fix a parameter, rather than marginalizing over it, perhaps because one wants to see what happens for particularly interesting values of that parameter. So for instance one may fix $\Omega_\Lambda^{(0)}$ to be $\Omega_\Lambda^{(0)} = 0$ and evaluate $L(\Omega_m^{(0)}, \Omega_\Lambda^{(0)} = 0)$. Then the result will obviously depend on the fixed value.

When that value is the maximum likelihood estimator, the likelihood is said to be *maximized* (as opposed to *marginalized*) with respect to that parameter.

If this is your first encounter with maximum likelihood methods, warm up by proving that if we have the Gaussian likelihood $f(x_i; \mu, \sigma^2)$

$$f(x_i; \mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp \left[-\frac{1}{2} \sum_i^n \frac{(x_i - \mu)^2}{\sigma^2} \right], \quad (13.17)$$

then the maximum likelihood estimator of μ is given by

$$\hat{\mu} = \frac{1}{n} \sum_i^n x_i. \quad (13.18)$$

Analogously, you can prove that the maximum likelihood variance estimator is

$$\hat{\sigma}^2 = \frac{1}{n} \sum_i^n (x_i - \hat{\mu})^2. \quad (13.19)$$

You may notice that this falls short of the standard result according to which the estimate of the sample variance has $(n - 1)$ instead of n at the denominator. In this case in fact the maximum likelihood estimator is biased, which means that its expectation value does not equal the “true” or “population” value. Indeed, maximum likelihood estimators are not necessarily unbiased although under some general conditions they are asymptotically (i.e. for $n \rightarrow \infty$) unbiased.

Let us conclude on a philosophical tone. One could say that the use of priors constitutes the whole difference between the Bayesian approach and the so-called *frequentist* one. The frequentist approach prefers not to deal with priors at all and therefore refuses to use Bayes’ theorem to convert theoretical parameters into random variables. Once a frequentist finds a maximum likelihood estimator (which as any other estimator is a function of data and therefore is a random variable), he or she tries to determine its distribution as a function of the assumed distribution of the data. In most cases, this is done by generating numerically many mock datasets and calculating for each dataset the estimator, deriving then its approximate distribution. This Monte Carlo approach is the hallmark of the frequentist approach. It is powerful, objective, and general but by rejecting priors fails to take into account previous knowledge. It is therefore suitable only when one can afford not to fully consider previous knowledge. This applies for instance when new experiments are much better than previous ones so that priors do not really matter and when each experiment measures only one parameter, say the mass of a particle, so that the outcome does not depend on other poorly measured

parameters. Both features characterize most particle physics experiments and this explains why most particle physicists are frequentist. Astrophysics and cosmology live in another experimental world: data are hard to come by, observations cannot be twisted and repeated as easily as in a laboratory, models are characterized by many correlated parameters, and every drop of previous information, even loosely related to a given parameter, has to be taken into account. Most of the evidence for dark energy comes from *combining* CMB and supernovae priors, each of them measuring many correlated parameters at once. It is no surprise that Bayesian methods are so popular in astrophysics and cosmology.

13.2 Model selection

So far we have been working within a given model. When we choose a model to test, we also select some free functions that define the model and that we parametrize in some convenient way. If we decide to change a model, e.g., from the uncoupled dark energy model with $w_{\text{DE}} = \text{constant}$ to a specific $f(R)$ model, we have to start a new process so that the likelihood will give us a new set of best fit parameters. But how do we decide whether the $f(R)$ model is better than the dark energy model with $w_{\text{DE}} = \text{constant}$?

This is a problem of *model selection*, rather than model optimization. One possibility (the *frequentist* approach) is to simply evaluate the “goodness of fit”: once we have the best fit parameters for models A and B, we calculate the χ^2 statistics of the model prediction with respect to data and choose the one with better χ^2 statistics (which is not necessarily the one with lowest χ^2 because the χ^2 statistics depends also on the number of degrees of freedom, namely on the number of independent data minus the number of free parameters). Beside the intrinsic problem of any frequentist approach (e.g., lack of priors), this is often too rough a guide to selection, mostly because if the model B includes a parameter that is poorly constrained by the data it would not help in the fit but it would still be counted as an extra degree of freedom and this would unfairly penalize it. Imagine for instance two very similar dark energy models, A and B, with two parameters each. Suppose that the model B predicts some peculiar feature at the redshift $z = 3$, e.g., cluster abundance, and that feature depends on a third parameter. The model B is interesting also because of this unique prediction but it would be unfairly penalized by current constraints, since we have very limited knowledge of high-redshift clusters so far. A χ^2 test would presumably conclude that the model A fits existing data as well as the model B but with one parameter less and therefore it would win.

To overcome this problem we can instead use another model selection procedure, called *evidence* or marginal likelihood. We have already seen in Eq. (13.6) that the

evidence is defined as the likelihood integral over the whole domain

$$E(\mathbf{x}; M) = \int f(\mathbf{x}; \theta_i^M) p(\theta_i^M) d^n \theta_i^M, \quad (13.20)$$

where as before $\mathbf{x} = (x_1, x_2, \dots)$ are random data, θ_i^M are n theoretical parameters that describe the model M , f is the likelihood function, and p is the prior probability of the parameter θ_i^M . Note that we have added a superscript M to remember that the parameters refer to some model M .

If we want to compare two models M_1 and M_2 , then we calculate the Bayes' ratio [543]

$$B_{12} = \frac{\int f(\mathbf{x}; \theta_i^{M_1}) p(\theta_i^{M_1}) d^n \theta_i^{M_1}}{\int f(\mathbf{x}; \theta_i^{M_2}) p(\theta_i^{M_2}) d^n \theta_i^{M_2}}. \quad (13.21)$$

A Bayes' ratio $B_{12} > 1$ (< 1) says that the current data favors the model M_1 (M_2). If we have any reason to weigh the models in some way, we can assign a model prior $p(M_j)$ and use Bayes' theorem again to write

$$L(M; \mathbf{x}) \propto E(\mathbf{x}; M) p(M), \quad (13.22)$$

and evaluate the ratio

$$\frac{L(M_1; \mathbf{x})}{L(M_2; \mathbf{x})} = B_{12} \frac{p(M_1)}{p(M_2)}. \quad (13.23)$$

Generally, however, one assumes that $p(M_1) = p(M_2)$.

Suppose now that a certain parameter θ_n is very poorly constrained by the data x_i . This implies that the likelihood $f(x_i; \theta_i)$ is practically independent of θ_n , that is, f remains almost constant when varying θ_n . Then if the prior is factorizable (which is often the case) so that $p(\theta_i) = \prod_i p_i(\theta_i)$, we see that the integral over θ_n decouples. Since the priors are just standard probability distribution functions we have $\int p_n(\theta_n) d\theta_n = 1$, so that as expected θ_n does not enter the evidence integral. The evidence therefore correctly discards poorly constrained parameters and does not penalize models for introducing them. The blame is where it belongs: poor data.

If the likelihood and the prior can both be approximated by Gaussian distributions, we can evaluate the evidence analytically. Let us assume then an uncorrelated Gaussian likelihood with best fit parameters $\theta_i^{(B)}$ and variances $\sigma_{B,i}$ and an uncorrelated Gaussian prior with means $\theta_i^{(P)}$ and variances $\sigma_{P,i}$. The posterior can be

written as

$$\begin{aligned}
 L(\theta_i) &= \prod_i f(\mathbf{x}; \theta_i) p(\theta_i) \\
 &= \prod_i f_{\max,i} (2\pi\sigma_{P,i}^2)^{-1/2} \exp \left[-\frac{(\theta_i - \theta_i^{(B)})^2}{2\sigma_{B,i}^2} - \frac{(\theta_i - \theta_i^{(P)})^2}{2\sigma_{P,i}^2} \right] \\
 &= \prod_i f_{\max,i} (2\pi\sigma_{P,i}^2)^{-1/2} \exp \left[-\frac{1}{2} \frac{(\theta_i - \theta_i^*)^2}{\sigma_{i*}^2} \right] \exp \left[-\frac{1}{2} \frac{(\theta_i^{(B)} - \theta_i^{(P)})^2}{\sigma_{B,i}^2 + \sigma_{P,i}^2} \right],
 \end{aligned} \tag{13.24}$$

where $f_{\max,i}$ is the i -th likelihood maximum and where the posterior mean and variance for each i are given, respectively, by

$$\theta_i^* = \frac{\sigma_{B,i}^2 \theta_i^{(P)} + \sigma_{P,i}^2 \theta_i^{(B)}}{\sigma_{B,i}^2 + \sigma_{P,i}^2}, \tag{13.25}$$

$$\sigma_{i*}^2 = \frac{\sigma_{P,i}^2 \sigma_{B,i}^2}{\sigma_{B,i}^2 + \sigma_{P,i}^2}. \tag{13.26}$$

The evidence is therefore

$$\begin{aligned}
 E &= \int f(\mathbf{x}; \theta_i) p(\theta_i) d\theta_i \\
 &= \prod_i f_{\max,i} \frac{\sigma_{i*}}{\sigma_{P,i}} \exp \left\{ -\frac{1}{2} \left[\left(\frac{\theta_i^{(B)}}{\sigma_{B,i}} \right)^2 + \left(\frac{\theta_i^{(P)}}{\sigma_{P,i}} \right)^2 - \left(\frac{\theta_i^*}{\sigma_{i*}} \right)^2 \right] \right\}.
 \end{aligned} \tag{13.27}$$

We see that the evidence is determined by three factors. The first, $f_{\max,i}$, is the likelihood maximum and expresses how well the model fits the data. The second is a ratio of parameter volumes: if we take the variance as a measure of the available parameter space for the i -th parameter, this factor expresses how the parameter volume changes from the prior to the posterior. Every factor $\sigma_{i*}/\sigma_{P,i} = \sigma_{B,i}/(\sigma_{B,i} + \sigma_{P,i})^{1/2}$ is smaller than unity, so adding more parameters penalizes the evidence, quantifying Ockham's razor argument.² If however the data do not constrain the i -th parameter, i.e. if $\sigma_{B,i} \gg \sigma_{P,i}$, then the i -th factor $\sigma_{i*}/\sigma_{P,i}$ is close to unity and there is no penalization. Finally the third factor penalizes the evidence if the best-fit i -th parameter or the prior mean differ appreciably from the posterior mean θ_i^* : although the new data might justify that parameter, the overall agreement

² William of Ockham (c.1288–c.1348), a Franciscan theologian, was known for his principle “*Entia non sunt multiplicanda sine necessitate*” (although this particular formulation is probably apocryphal) – one should not multiply entities beyond necessity.

including the prior does not seem to require it. Here again, if data constraints are very weak (large $\sigma_{B,i}$) then there is no penalization. You can work out another example in problem 13.3.

It is a matter of straightforward algebra to extend the expression to correlated Gaussian parameters. If the evidence integral is

$$E = \int f(\mathbf{x}; \theta_i) p(\theta_i) d\theta_i$$

$$\approx f_{\max} \int \exp \left[-\frac{1}{2}(\theta_i - \theta_i^{(B)}) L_{ij} (\theta_j - \theta_j^{(B)}) - \frac{1}{2}(\theta_i - \theta_i^{(P)}) P_{ij} (\theta_j - \theta_j^{(P)}) \right] d\theta_i, \quad (13.28)$$

where $\theta_i^{(B)}$ are the best fit estimators, $\theta_i^{(P)}$ are the prior means, L_{ij} in the exponential factor is the inverse of the covariance matrix of the likelihood (or Fisher matrix, see the next section) and P_{ij} is the inverse of the covariance matrix of the prior, we obtain

$$E = f_{\max} \frac{|\mathbf{F}|^{-1/2}}{|\mathbf{P}|^{-1/2}} \exp \left[-\frac{1}{2}(\theta_i^{(B)} L_{ij} \theta_j^{(B)} + \theta_i^{(P)} P_{ij} \theta_j^{(P)} - \tilde{\theta}_i F_{ij} \tilde{\theta}_j) \right], \quad (13.29)$$

where $\mathbf{F} = \mathbf{P} + \mathbf{L}$ and $\tilde{\theta}_i = (\mathbf{F}^{-1})_{im} [L_{mj} \theta_j^{(B)} + P_{mj} \theta_j^{(P)}]$.

The evidence is often not easy to evaluate because it requires a multidimensional integration over the whole parameter space. Several approximation or alternative model selection techniques have been proposed (see for instance the excellent review [544]). They are however only justified in specific cases and may give conflicting results, sometimes leading to controversies [545, 546]. Whenever possible, the evidence integral should be used instead.

Let us now come back to the Bayes' factors, i.e. the ratio of the evidences. Once we have calculated this ratio we are still to decide how to gauge it in favor of the model A or B. There is no absolute way to achieve this: large or small factors should incline us towards one of the two models over the other one, but there is no absolute "statistics" to associate to any specific level. The scale most used in literature is called Jeffrey's scale. If $|\ln B_{12}| < 1$ there is no evidence in favor of any of the models ("inconclusive evidence"); if $|\ln B_{12}| > 1$ there is a "weak evidence"; $|\ln B_{12}| > 2.5$ means "moderate evidence"; $|\ln B_{12}| > 5$ means "strong evidence." Of course this terminology is purely suggestive and not to be taken literally. We can consider it as a practical bookkeeping device. When the data promote a model from weakly to moderately to strongly "evident," it is time to take it seriously and challenge aggressively.

13.3 Fisher matrix

As straightforward and versatile as the likelihood method is, it is still often too complicated or computing-expensive to implement, especially when there are more than a few parameters involved. In fact there are some cases in which several tens or hundreds of parameters are present.

One could think that a model with more than 3 or 4 free parameters does not deserve the name of model and even less that one of “theory”. However every theory begins by representing a vast dataset with a smaller set of numbers. And since cosmological experiments may easily collect terabytes of data, reducing them to 10, 100, or 1000 numbers should be seen already as a great progress towards a unified description (if there is one!).

Anyway, the problem with the likelihood is that we need to evaluate $L(\theta_i)$ for every θ_i , or at least for *many* θ_i , e.g., for a grid of, say, 10 values for each dimension in parameters space. If there are 10 parameters, this means 10^{10} different evaluations. If each evaluation takes a second (say, a run of a CMB code), we are in for a waiting time of 300 years . . .

One way out is to use a Monte Carlo approach. Instead of building a full grid, one explores the landscape with random jumps. The size of the jumps in turn may be related to the steepness of the function (smaller jumps over rough terrain, larger ones over flatlands). This technique will grow with the number D of dimensions (parameters) as D , instead of exponentially as in full grid method. But this might still be a lot: a typical Markov chain exploration can take hundred of thousands of computations.

It is time to think of something faster: the Fisher matrix. The idea is straightforward: to approximate the full likelihood with a (multivariate) Gaussian distribution,

$$L \approx N \exp \left[-\frac{1}{2} (\theta_i - \hat{\theta}_i) F_{ij} (\theta_j - \hat{\theta}_j) \right], \quad (13.30)$$

where the values $\hat{\theta}_i$, the maximum likelihood estimators, are functions of the data, and F_{ij} , the Fisher (or information) matrix, is the inverse of the correlation matrix. It is crucial to pay attention to the fact that the likelihood is a Gaussian function of the *parameters*, not (or not only) of the data. We often assumed in the previous sections the data to be Gaussian but never, so far, did the same for the parameters. The form (13.30) is of course a crude approximation. One could hope however that it is a reasonable approximation at least near the peak of the distribution, given that around a local maximum every smooth function (in this case $\ln L$) can be approximated as a quadratic function. Therefore we expect this approximation to work better for θ_i close to their estimators $\hat{\theta}_i$.

Expanding the exponent of a generic likelihood near its peak (i.e. near the maximum likelihood (ML) value $\hat{\theta}_i$ of the parameters) as

$$\ln L(\theta_i) \approx \ln L(\hat{\theta}_i) + \frac{1}{2} \frac{\partial^2 \ln L(\theta_i)}{\partial \theta_i \partial \theta_j} \bigg|_{\text{ML}} (\theta_i - \hat{\theta}_i)(\theta_j - \hat{\theta}_j) \quad (13.31)$$

(naturally the first derivatives are absent because they vanish at the peak) we find, comparing with Eq. (13.30), that the normalization $N = L(\hat{\theta}_i)$ depends only on the data and that the *Fisher matrix* (FM) is defined as

$$F_{ij} \equiv - \frac{\partial^2 \ln L(\theta)}{\partial \theta_i \partial \theta_j} \bigg|_{\text{ML}}. \quad (13.32)$$

Before proceeding further, let us remark that actually the FM is defined as the *expected* value of the matrix $-\partial^2 \ln L / \partial \theta_i \partial \theta_j$, to be obtained by averaging the matrix over the data distribution, i.e.

$$F_{ij} \equiv - \left\langle \frac{\partial^2 \ln L(\theta)}{\partial \theta_i \partial \theta_j} \right\rangle = - \int \frac{\partial^2 \ln L(\theta)}{\partial \theta_i \partial \theta_j} L(\mathbf{x}; \theta) d\mathbf{x}. \quad (13.33)$$

However, within the approximation (13.30), the two definitions coincide.

You may say now that in order to find the ML estimator we still have to build the full likelihood: does this again require the 10^{10} evaluations of $L(\theta_i)$ that we mentioned above? Well, we could answer that there are fast numerical methods to search for maxima in a multidimensional function without spanning the whole parameter space. For instance, in one dimension, if we can guess that the parameter is near $\theta^{(0)}$ then we can expand the derivative of the log-likelihood $\mathcal{L} = -\ln L$ as follows

$$\mathcal{L}_{,\theta}(\theta) \approx \mathcal{L}_{,\theta}(\theta^{(0)}) + \mathcal{L}_{,\theta\theta}(\theta - \theta^{(0)}), \quad (13.34)$$

and estimate the minimum of $\mathcal{L}$ (i.e. the maximum of L) by putting $\mathcal{L}_{,\theta}(\theta) = 0$. Then we find the approximation

$$\theta^{(1)} = \theta^{(0)} - \frac{\mathcal{L}_{,\theta}}{\mathcal{L}_{,\theta\theta}} \bigg|_{\theta^{(0)}}, \quad (13.35)$$

which could be iterated by assuming a new guess $\theta^{(1)}$ instead of $\theta^{(0)}$. This method, called Newton–Raphson, is extremely fast for well-behaved likelihood functions and can be directly generalized to the multidimensional case. However perhaps the most useful application of the Fisher formalism is to the cases in which we do not need to search for the likelihood peak because we already know from the start the ML estimator: when we are *simulating* an experiment.

Suppose we want to forecast how well a future supernovae experiment, which is supposed to collect $n = 10\,000$ supernovae light curves and to derive their peak

magnitude m_i with errors σ_i , is capable of constraining the cosmological parameters $\Omega_m^{(0)}, \Omega_\Lambda^{(0)}$. Let us start by assuming that the n random variables $m_i(z_i)$ follow a PDF with known variance σ_i and mean $m_{\text{th}}(z_i; \Omega_m^{(0)}, \Omega_\Lambda^{(0)}) = 5 \log_{10} d_L(z_i; \Omega_m^{(0)}, \Omega_\Lambda^{(0)}) + C$. Here we take the PDF to be Gaussian but we could also assume another PDF if we think it describes the data. Since the data PDF is assumed to be Gaussian we can immediately form the likelihood (neglecting the normalization constant):

$$L_m \approx \exp \left[-\frac{1}{2} \sum_i \frac{(m_i - m_{\text{th}}(z_i))^2}{\sigma_i^2} \right] = \exp \left(-\frac{1}{2} \mu_i C_{ij}^{-1} \mu_j \right). \quad (13.36)$$

Here we have expressed the argument of the exponential in a slightly more general way: we have introduced the vector $\mu_i \equiv m_i - m_{\text{th}}(z_i)$ and the correlation matrix C_{ij} , that in this particular case is rather trivial

$$C = \text{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2 \dots). \quad (13.37)$$

When we discuss dark energy, we are interested in the parameters such as $\Omega_m^{(0)}, \Omega_\Lambda^{(0)}$. So we wish to produce a likelihood function of $\Omega_m^{(0)}, \Omega_\Lambda^{(0)}$, something in the form of Eq. (13.30) like

$$L(\Omega_m^{(0)}, \Omega_\Lambda^{(0)}) = \exp \left[-\frac{1}{2} (\Omega_i^{(0)} - \hat{\Omega}_i^{(0)}) F_{ij} (\Omega_j^{(0)} - \hat{\Omega}_j^{(0)}) \right], \quad (13.38)$$

where F_{ij} is of course our Fisher matrix and i, j run over the subscripts m, Λ . Since real data are not yet present, we do not have the ML estimators $\hat{\Omega}_i^{(0)}$. However we are simulating the future experiment, so we may take for estimators the values $m_{\text{th}}(z_i; \Omega_m^{(0)F}, \Omega_\Lambda^{(0)F})$ obtained using some fiducial cosmology $\Omega_m^{(0)F}, \Omega_\Lambda^{(0)F}$, for instance $\Omega_m^{(0)F} = 0.3, \Omega_\Lambda^{(0)F} = 0.7$. This means that we will find the confidence regions only around this particular parameter set. If we decide to change fiducial values, we have to redo our calculations and all our results will change in some way.

The Fisher matrix of the likelihood (13.36) is then

$$F_{ij} = - \left. \frac{\partial^2 \ln L_m}{\partial \Omega_i^{(0)} \partial \Omega_j^{(0)}} \right|_F = \sum_n \frac{1}{\sigma_n^2} \left. \frac{\partial^2 m_{\text{th}}(z_n; \Omega_m^{(0)}, \Omega_\Lambda^{(0)})}{\partial \Omega_i^{(0)} \partial \Omega_j^{(0)}} \right|_F. \quad (13.39)$$

Notice that F_{ij} is not diagonal even if the original correlation matrix C_{ij} was. Since the same $\Omega_m^{(0)}, \Omega_\Lambda^{(0)}$ appear in all $m_{\text{th}}(z_n)$, we vary the likelihood of obtaining *all* m_i by varying $\Omega_{m,\Lambda}^{(0)}$. We can now use Eq. (13.38) to derive the confidence errors for $\Omega_m^{(0)}, \Omega_\Lambda^{(0)}$. In practice, what we have developed so far is a formalism to propagate the errors from the observational errors σ_i to the cosmological parameters. The errors σ_i , in turn, must be based on the expected performance of the experiment

and often their derivation is the most complicated step, involving many fine details of the observations. Calculating numerically the second-order partial derivatives in the Fisher matrix requires only a few estimations of the likelihood for each of the parameters; if we have 10 parameters this makes a few tens of calculations instead of the 10^{10} we mentioned at the beginning of this section.

Once we have reduced our likelihood into a Gaussian form, the Fisher matrix is all we need to derive all the properties. The rest of this section is concerned with various ways to manipulate the Fisher matrix to achieve several results.

Suppose we decide to switch from a set of parameters x_i to another one $y_j(x_i)$, for instance from $\Omega_m^{(0)}, \Omega_\Lambda^{(0)}$ to the spatial curvature $\Omega_K^{(0)} = 1 - \Omega_m^{(0)} - \Omega_\Lambda^{(0)}$ and their ratio $R_{m\Lambda} = \Omega_m^{(0)}/\Omega_\Lambda^{(0)}$. If we know the Fisher matrix for x_i , the approximate likelihood is

$$L = \exp\left(-\frac{1}{2}\tilde{x}_i F_{ij}^{(x)} \tilde{x}_j\right), \quad (13.40)$$

where $\tilde{x}_i = x_i - x_i^{\text{ML}}$. Approximating y_j near x_i^{ML} as

$$y_j \approx y_j^{\text{ML}} + \left.\frac{\partial y_j}{\partial x_i}\right|_{\text{ML}} (x_i - x_i^{\text{ML}}), \quad (13.41)$$

where $y_j^{\text{ML}} \equiv y_j(x^{\text{ML}})$, we can write

$$\tilde{y}_j \equiv y_j - y_j^{\text{ML}} = J_{ji}^{-1} \tilde{x}_i. \quad (13.42)$$

Here $J_{ji} \equiv (\partial x_j / \partial y_i)_{\text{ML}}$ is the transformation Jacobian evaluated on the ML estimators. Then we have

$$\tilde{x}_i = J_{i\ell} \tilde{y}_\ell, \quad (13.43)$$

and we can find the new Fisher matrix by substituting into Eq. (13.40) simply as

$$F_{\ell m}^{(y)} = J_{i\ell} F_{ij}^{(x)} J_{jm}, \quad (13.44)$$

which is summed over indices. We can say that the Fisher matrix transforms as a tensor. Notice that the Jacobian matrix does not need to be a square matrix. The old parameters x_j can be projected in fact onto a smaller number of new parameters y_i .

One may wonder why the Jacobian does not enter also in the transformation from the volume element $dx_1 dx_2 \dots$ to the new element $dy_1 dy_2 \dots$, so that $L(y_j) = |J| L[x_i(y_j)]$. This would imply an additional logarithmic term $\ln |J|$ in the transformed probability function, spoiling the Gaussian approximation altogether. However near the ML values we can approximate $|J|$ with $|J_{\text{ML}}|$ and include this constant factor in the overall normalization. That is, forget about it.

Let us apply the transformation technique to an interesting problem. We have used extensively the parametrization around $a_0 = 1$ of the equation of state

$w_{\text{DE}}(a) = w_0 + w_1(1 - a)$ [Eq. (7.93)]. We could however have expanded $w_{\text{DE}}(a)$ around any other point a_p and write instead $w_{\text{DE}}(a) = w_p + w_1(a_p - a)$, where

$$w_p = w_0 + w_1(1 - a_p). \quad (13.45)$$

We can now ask the question whether the constraint we obtain on w_p (i.e. $\sigma_{w_0}^2$) is tighter than the one on w_0 , that is whether we can better rule out say $w_{\text{DE}} = -1$ at a_p than at a_0 . The problem consists therefore in finding the value a_p (called *pivot point*) that minimizes the variance of $w_{\text{DE}}(a)$. Denoting the maximum likelihood estimators (or fiducial values) with $\hat{w}_0, \hat{w}_1$, this occurs for the value of a which is the solution of the following equation,

$$\begin{aligned} \frac{d}{da} [\langle [(w_0 - \hat{w}_0) + (1 - a)(w_1 - \hat{w}_1)]^2 \rangle] \\ = \frac{d}{da} [\sigma_{w_0}^2 + (1 - a)^2 \sigma_{w_1}^2 + 2(1 - a)\rho\sigma_{w_0}\sigma_{w_1}] \\ = -2(1 - a)\sigma_{w_1}^2 - 2\rho\sigma_{w_0}\sigma_{w_1} = 0. \end{aligned} \quad (13.46)$$

Here $\sigma_{w_i}^2 \equiv \langle (w_i - \hat{w}_i)^2 \rangle$ for $i = 0, 1$ and $\rho \equiv \langle (w_0 - \hat{w}_0)(w_1 - \hat{w}_1) \rangle / (\sigma_{w_0}\sigma_{w_1})$ is the correlation coefficient. Then we obtain [547]

$$a_p = 1 + \frac{\rho\sigma_{w_0}}{\sigma_{w_1}}. \quad (13.47)$$

In terms of the two-dimensional Fisher matrix F_{ij} for w_0, w_1 , we can write

$$\sigma_{w_0}^2 = (\mathbf{F}^{-1})_{11}, \quad \sigma_{w_1}^2 = (\mathbf{F}^{-1})_{22}, \quad \rho\sigma_{w_0}\sigma_{w_1} = (\mathbf{F}^{-1})_{12}. \quad (13.48)$$

The transformation from $\mathbf{p} = (w_0, w_1)$ to $\mathbf{q} = (w_p, w_1)$ is achieved by using Eq. (13.44) with the transformation matrix

$$\mathbf{J} = \frac{\partial \mathbf{p}}{\partial \mathbf{q}} = \begin{pmatrix} 1 & 1 - a_p \\ 0 & 1 \end{pmatrix}. \quad (13.49)$$

It is straightforward to verify that with this transformation the new matrix $\mathbf{F}_p = \mathbf{J}^t \mathbf{F} \mathbf{J}$ is diagonal (the superscript t denotes transpose) and its inverse is:

$$\mathbf{F}_p^{-1} = \begin{pmatrix} \sigma_{w_0}^2(1 - \rho^2) & 0 \\ 0 & \sigma_{w_1}^2 \end{pmatrix}. \quad (13.50)$$

The parameters w_p, w_1 are therefore uncorrelated. Moreover, as expected, the error on $w_p, \sigma_{w_p}^2 \equiv \sigma_{w_0}^2(1 - \rho^2)$, is always smaller than $\sigma_{w_0}^2$.

What if we want to *maximize* the likelihood with respect to some parameter? This means, if you remember, to fix one of the parameters to its maximum likelihood estimator. With the Fisher matrix this is really trivial, since fixing a parameter to its maximum likelihood estimator means putting the difference $\theta_i - \hat{\theta}_i = 0$ and

therefore to discard all entries in the Fisher matrix related to the i -th parameter. In practice, this means that one removes from the Fisher matrix the rows and columns of the maximized parameters.

What about *marginalization* then? Take a general two-dimensional Gaussian PDF

$$G(x_1, x_2) = N \exp \left[-\frac{1}{2(1-\rho^2)} \left(\frac{x_1^2}{\sigma_1^2} + \frac{x_2^2}{\sigma_2^2} - 2\frac{\rho x_1 x_2}{\sigma_1 \sigma_2} \right) \right], \quad (13.51)$$

where ρ is the correlation factor. This PDF can be written as

$$G(X_i) = N \exp \left[-\frac{1}{2} (X_i C_{ij}^{-1} X_j) \right], \quad (13.52)$$

where $X_i \equiv x_i - \mu_i$ (generalizing to non-zero μ 's), and

$$C = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}. \quad (13.53)$$

Let us now evaluate the integral $\int G(x_1, x_2) dx_2$ over the whole real domain. The result is given by

$$G(x_1) = \tilde{N} \exp[-x_1^2/(2\sigma_1^2)], \quad (13.54)$$

where $\tilde{N}$ is a new normalization constant. The new correlation “matrix” is now simply $C_{11} = \sigma_1^2$.

In terms of the Fisher matrix $F = C^{-1}$ we see that the outcome of the marginalization has been the removal from $F^{-1} = C$ of the rows and columns related to the second parameter. This trick remains true for any number of dimensions: to marginalize over the j -th parameter, one simply needs to remove from the *inverse* of the Fisher matrix F^{-1} the j -th row and column; to marginalize at once over several parameters, one removes all the rows and columns related to those parameters. As a consequence, the diagonal of the inverse Fisher matrix contains the *fully marginalized* 1σ errors of the corresponding parameters (i.e. the errors one gets on the i -th parameter after marginalizing over all the others)

$$\sigma_i^2 = (F^{-1})_{ii}. \quad (13.55)$$

This latter property is probably the most useful and time-saving feature of the whole Fisher method. Be warned however that the procedure of inverting and striking out rows and columns is in general numerically unstable if the matrix contains small eigenvalues. There are more stable algorithms that perform this operation [547].

Often we want to reduce the Fisher matrix to a 2×2 matrix F_2 for two parameters, say θ_1, θ_2 , because then it is easy to plot the resulting two-dimensional confidence regions, defined as the regions of constant likelihood that contain a

predetermined fraction of the total likelihood volume. Since the problem has been reduced from the start to Gaussianity, we will necessarily have ellipsoidal confidence regions on the plane θ_1, θ_2 . Looking at the form of the two-dimensional Gaussian PDF (13.51), you will realize that the semiaxes of the ellipses are oriented along the eigenvectors of $\mathbf{F}_2^{-1}$, that is, they form an angle

$$\tan 2\alpha = \frac{2\rho\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2}, \quad (13.56)$$

with the coordinate axes. Moreover, the semiaxes ratio is equal to the square root of the eigenvalues ratio. The length of the semiaxes depends clearly on the level of confidence. If we take the semiaxes length along the i -th eigenvector equal to $\sqrt{\lambda_i}$, where λ_i is the i -th eigenvalue, we are finding the 1σ region, but because we are in two dimensions, this level does not contain 68.3% of the probability but rather less than 40%. Instead, we find by integrating a two-dimensional Gaussian that the *one-dimensional* “ 1σ ” region corresponding to 68.3% of probability content is found for semiaxes which are roughly 1.51 times the eigenvalues. Regions at 95.4% and 99.7% correspond to semiaxes 2.49 and 3.44 times the eigenvalues, respectively. The area of the 68.3% ellipses is πab , if a and b are the semiaxes length, that is 1.51 times the eigenvalues. The area is therefore equal to $(1.51)^2\pi(\det \mathbf{F}_2)^{-1/2}$. Since an experiment is more constraining when the confidence region is smaller, one can define a simple but useful figure of merit (FOM) as [547]

$$\text{FOM} = \sqrt{\det \mathbf{F}_2}. \quad (13.57)$$

Notice however that the FOM is often defined to be the area at 95%, or some other similar but not equivalent choice.

The FOM is particularly relevant to dark energy parameters such as w_0, w_1 [see, e.g., Eq. (7.93)]. The FOM naturally depends on how many parameters have been marginalized. Every parameter marginalization increases (or more exactly, does not reduce) the amount of uncertainty with respect to a maximized likelihood and therefore decreases the available information and the FOM of the final set of parameters, as we show in Fig. 13.1.

All these simple rules are really good news for practical work. The bad news comes when they do not work. The major problem, in practice, is when the Fisher matrix itself is singular. Then there is no inverse and no marginalization. But the Fisher matrix can be singular only when rows or columns are not linearly independent. It is easy to see when this happens. If $L(\theta_1, \theta_2)$ depends on the two parameters through a constant combination, e.g., $a\theta_1 + b\theta_2$, then the Fisher matrix will be singular.

Let us turn this bug into a feature. If the Fisher matrix is singular, then it means that there is a linear combination of two or more parameters hidden somewhere

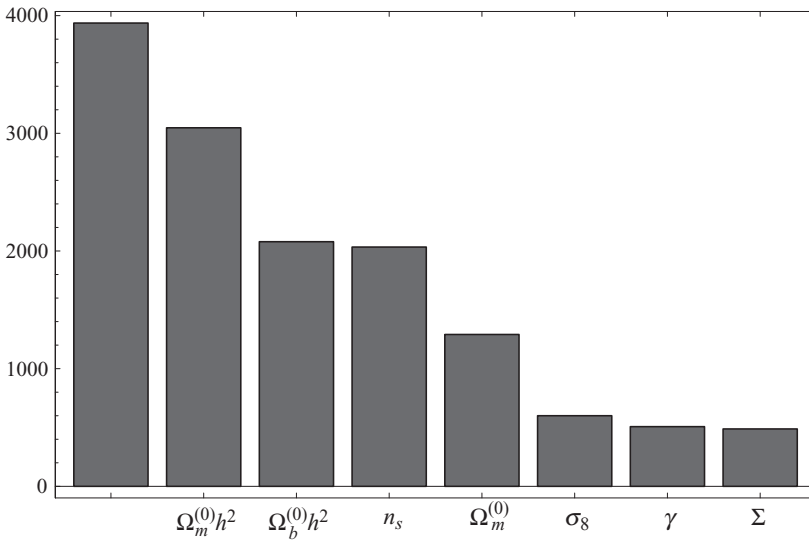


Figure 13.1 The first bar on the left is the FOM for w_0, w_1 with the parametrization (7.93) in a weak lensing experiment assuming all the other cosmological parameters have been fixed to their fiducial values. When additional parameters (listed on the bottom of the histogram) are marginalized, instead of being held fixed, the FOM reduces. From Ref. [497].

in the likelihood. Therefore, we can substitute a new parameter $\hat{\theta}$ in place of that combination, e.g., $\hat{\theta} = a\theta_1 + b\theta_2$ and remove the singularity by restricting ourselves to $\hat{\theta}$ instead of the original pair. Actually we should have done this from the start, since if the physics depends only on the combination $a\theta_1 + b\theta_2$ there is no way we can distinguish between θ_1, θ_2 . It is only this combination that matters and we should replace it by $\hat{\theta}$. We say in this case that there is a *degeneracy* between θ_1 and θ_2 . Sometimes, however, it is not obvious at all that this was the case and the singularity of the Fisher matrix is a warning for us to look harder.

The only real problem is when there is *almost* a singularity. If the combination is given by $a\theta_1 + b\theta_2 + cf(z)\theta_1^2$, then there should be no singularity because of the non-constant term (we are thinking here of observations at several z 's). However, if a, b are of the order of unity while $c = 10^{-10}$, then there is a high degree of degeneracy, albeit not a total one. In this case the Fisher matrix may behave in a dangerous way, with extremely small eigenvalues and unstable inversions. This is the case that requires a human brain. It is our duty to understand the physical cause of this quasi-degeneracy and redefine the parameters, perhaps giving up the possibility of discriminating between θ_1, θ_2 and focusing on the combined term $\hat{\theta} = a\theta_1 + b\theta_2 + cf(z)\theta_1^2$. Or we may find additional priors (e.g., other experiments)

that give separate information on one of the quasi-degenerate parameters and break the degeneracy.

This brings us to another advantage of the Fisher matrix approach. How do we add priors to a Fisher matrix F_{ij} ? If the prior is the outcome of another experiment and we have the Fisher matrix $F_{ij}^{(p)}$ of that experiment, then the problem reduces to multiplying a Gaussian likelihood by another Gaussian likelihood, obtaining a new Gaussian likelihood. If the experiments have the same ML estimators or the same fiducial model, as in the case in which we simulate them, the new Fisher matrix is given by

$$F_{ij}^{(\text{tot})} = F_{ij} + F_{ij}^{(p)}. \quad (13.58)$$

As simple as this: combining the information from two forecasts (with the same fiducial model) means summing their Fisher matrices. In so doing one has to ensure that the parameters and their order are exactly the same for both matrices: trivial, but a most likely source of practical confusion. If one of the experiments constrains only a subset of the total parameters (for instance, supernovae experiments do not constrain the primordial perturbation slope n_s), it means that it contains no information on that subset, and therefore the corresponding rows and columns are to be put to zero. This means that the two Fisher matrices are rendered of the same rank by filling the one with less parameters (say $F^{(p)}$) with zeros in the correct position. For instance if we only want to add the information that the single m -th parameter comes with an error σ_m then we add the Fisher matrix (no sum on m)

$$F_{ij}^{(p)} = \frac{\delta_i^m \delta_j^m}{\sigma_m^2}. \quad (13.59)$$

So you see that in this case $F^{(p)}$ would be utterly singular but the total $F^{(\text{tot})}$ is not (unless of course F was singular as well for the same parameter, bad luck really).

Let us mention the final point about the Fisher matrix. A statistical theorem known as Cramer–Rao inequality states that the minimal variance of an unbiased estimator cannot be less than $(F^{-1})_{ii}$ (which means first to take the inverse and *then* take the i -th term on the diagonal). In this sense the Fisher matrix gives the minimal error one can hope to achieve. If you are very optimistic then the Fisher matrix is your tool. Notice, however, that the maximum likelihood estimators need not be unbiased estimators at all, although they are unbiased for large samples (asymptotically unbiased) otherwise they would be of little utility. So we could end up in producing the best possible error estimate for some unbiased estimators which we do not know how to determine!

Once we accept the Gaussian approximation, the Fisher matrix embodies all the information we have on the problem. The manipulation of the Fisher

matrix therefore is all we need. To recapitulate, there are five golden rules of *fisherology*:

1. To *transform* variables, multiply the Fisher matrix on the right and on the left by the transformation Jacobian.
2. To *maximize* over some parameters, remove from the matrix the rows and the columns related to those parameters.
3. To *marginalize* over some parameters, remove from the *inverse* matrix the rows and the columns related to those parameters (being careful about the numerical instability pointed out above).
4. To *combine* Fisher matrices from independent experiments with the same fiducial, sum the corresponding Fisher matrices, ensuring the same order of parameters, and, if necessary, inserting rows and columns of zeros for unconstrained parameters.
5. The *ellipsoidal confidence regions* have semiaxes lengths equal to the square root of the eigenvalues of the *inverse* Fisher matrix, while the semiaxes are oriented along the corresponding eigenvectors. The *area* of the ellipse (or volume of ellipsoid) is proportional to the square root of the determinant of the inverse Fisher matrix. The determinant of the Fisher matrix is an indicator of performance or a figure of merit.

If one wishes, one could define a new set of parameters by diagonalizing the Fisher matrix, obtaining (by an axes rescaling) circular (or spherical) confidence regions. In some cases this is useful because it reveals hidden properties (see Section 13.5). There are other cases in which the new parameters are so remote from any physical direct meaning that the exercise is futile. Notice that the confidence region volume (and therefore the FOM) does not change under the diagonalization.

13.4 The Fisher matrix for the power spectrum

Now we have all the tools to derive a very useful result, the Fisher matrix for an experiment that measures the galaxy power spectrum.

Suppose a future experiment will provide us with the Fourier coefficients $\delta_{\mathbf{k}}$ of a galaxy distribution and their power spectrum calculated for a set of m wavenumbers $\mathbf{k}_i$ in some redshift bin $z, z + \Delta z$. Our theory predicts the spectrum $P(k, z; p_i)$ as a function of, say, $p_i \equiv \Omega_m^{(0)}, \Omega_b^{(0)}, h, n_s$ etc. In any real survey with a galaxy density $n(z)$, however, the power spectrum will include the Poisson noise part that we estimated in Eq. (3.34):

$$\Delta_{\mathbf{k}}^2 \equiv \langle \delta_{\mathbf{k}} \delta_{\mathbf{k}}^* \rangle = \langle \delta_{\mathbf{k}} \delta_{-\mathbf{k}} \rangle = P(\mathbf{k}, z) + \frac{1}{n}. \quad (13.60)$$

Since the average galaxy density is estimated from the survey itself we have by construction $\langle \delta(x) \rangle = 0$ and therefore $\langle \delta_{\mathbf{k}_i} \rangle = 0$ for any $\mathbf{k}_i$. The coefficients $\delta_{\mathbf{k}_i}$ are complex variables in which the real and imaginary parts obey the same Gaussian

statistics. So now we calculate the Fisher matrix for only, say, the real parts of δ_{k_i} and the Fisher matrix for the whole δ_{k_i} is simply the sum of two identical Fisher matrices, i.e. twice the result for the real parts. However when we count the total number of independent modes we have to remember that only half of them are statistically independent since $\delta_k^* = \delta_{-k}$ so in fact we should finally divide by two the final result. That is, we can forget both factors.

If we assume the galaxy distribution to be well approximated by a Gaussian we can write the likelihood:

$$L = \frac{1}{(2\pi)^{m/2} \prod_i \Delta_i} \exp \left[-\frac{1}{2} \sum_i^m \frac{\delta_i^2}{\Delta_i^2} \right], \quad (13.61)$$

(where to simplify notation we write $\Delta_i = \Delta_{k_i}$, $\delta_i = \text{Re } \delta_{k_i}$) assuming that the measures at every k_i are statistically independent. When we simulate a future experiment, $P(k, z)$ is taken to be the theoretical spectrum of our fiducial model described by the parameters $p_j^{(F)}$. Then we have

$$\mathcal{L} = -\ln L = \frac{m}{2} \ln(2\pi) + \sum_i \ln \Delta_i + \sum_i \frac{\delta_i^2}{2\Delta_i^2}. \quad (13.62)$$

We further simplify the notation by suppressing the index i running over the k bins from Δ_i , δ_i and denote the differentiation with respect to the j -th parameter as $\Delta_{,j}$. Now from Eq. (13.32) the Fisher matrix for a particular z bin is

$$\begin{aligned} F_{\ell m} &= \left\langle \frac{\partial^2 \mathcal{L}}{\partial p_\ell \partial p_m} \right\rangle = \sum \left[\frac{\Delta_{,\ell m}}{\Delta} - \frac{\Delta_{,\ell} \Delta_{,m}}{\Delta^2} - \langle \delta^2 \rangle \left(\frac{\Delta_{,\ell m}}{\Delta^3} - 3 \frac{\Delta_{,\ell} \Delta_{,m}}{\Delta^4} \right) \right] \\ &= \frac{1}{2} \sum_i \frac{\partial \ln P_i}{\partial p_\ell} \frac{\partial \ln P_i}{\partial p_m} \left(\frac{n P_i}{1 + n P_i} \right)^2, \end{aligned} \quad (13.63)$$

[where we used $\langle \delta^2 \rangle = \Delta^2$ from Eq. (13.60)] calculated on the fiducial model.

For a more compact expression we can now approximate the sum with an integral over k . To do this we need to count how many modes lie in the bin defined by the modulus interval $k, k + dk$ and cosine interval $d\mu$, i.e. in the Fourier volume $2\pi k^2 dk d\mu$. The number of modes we can really use is limited by two factors: the size of the volume and the shot noise. Modes larger than the survey volume cannot be measured. Short modes sampled by only a few galaxies cannot be reliably measured either.

To take into account these limitations we discretize the Fourier space into cells of volume $V_{\text{cell}} = (2\pi)^3 / V_{\text{survey}}$, so that we have $2\pi k^2 dk d\mu / V_{\text{cell}} = (2\pi)^{-2} V_{\text{survey}} k^2 dk d\mu$ modes in the survey volume. The integral form of the Fisher

matrix is therefore given by [548, 549]

$$F_{\ell m} = \frac{1}{8\pi^2} \int_{-1}^{+1} d\mu \int_{k_{\min}}^{k_{\max}} k^2 dk \frac{\partial \ln P(k, \mu)}{\partial p_\ell} \frac{\partial \ln P(k, \mu)}{\partial p_m} \left[\frac{nP(k, \mu)}{nP(k, \mu) + 1} \right]^2 V_{\text{survey}}. \quad (13.64)$$

The factor

$$V_{\text{eff}} = \left[\frac{nP(k, \mu)}{nP(k, \mu) + 1} \right]^2 V_{\text{survey}} \quad (13.65)$$

can be seen as an effective survey volume. When $nP \gg 1$ the sampling is good enough to derive all the cosmological information that can be extracted from the survey and there is no need of more sources. For $nP \ll 1$ the effective volume is severely reduced. If we subdivide the data into several z independent bins, we can simply sum the Fisher matrices for every bin.

It is straightforward to extend the Fisher matrix calculation to a more general likelihood with full correlation. Consider a set of n Gaussian data $\mathbf{x}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$ distributed according to the likelihood

$$L = \frac{1}{(2\pi)^{n/2} \sqrt{\det \mathbf{C}}} \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^t \mathbf{C}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right], \quad (13.66)$$

where t denotes the transpose. We define the data matrix $\mathbf{D} = (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^t$. Then the covariance matrix is defined in all generality as the expected value of $\mathbf{D}$:

$$\langle \mathbf{D} \rangle = \mathbf{C}. \quad (13.67)$$

We can write, up to a constant

$$\mathcal{L} = -\ln L = \frac{1}{2} [\ln \det \mathbf{C} + \text{Tr} \mathbf{C}^{-1} \mathbf{D}] = \frac{1}{2} \text{Tr} [\ln \mathbf{C} + \mathbf{C}^{-1} \mathbf{D}], \quad (13.68)$$

where we used the matrix identity: $\ln \det \mathbf{C} = \text{Tr} \ln \mathbf{C}$. We suppose now that the theoretical parameters $\boldsymbol{\theta}$ are *both* in $\boldsymbol{\mu}$ and in $\mathbf{C}$. The Fisher matrix is then the expected value

$$F_{ij} = \left\langle \frac{\partial^2 \mathcal{L}}{\partial \theta_i \partial \theta_j} \right\rangle \equiv \langle \mathcal{L}_{,ij} \rangle. \quad (13.69)$$

To calculate $\langle \mathcal{L}_{,ij} \rangle$ we use the fact that for Gaussian data $\langle \mathbf{x} \rangle = \boldsymbol{\mu}$, and consequently

$$\langle \mathbf{D}_{,i} \rangle = 0, \quad \langle \mathbf{D}_{,ij} \rangle = \boldsymbol{\mu}_{,i} \boldsymbol{\mu}_{,j}^t + \boldsymbol{\mu}_{,j}^t \boldsymbol{\mu}_{,i}. \quad (13.70)$$

Notice that $\langle \mathbf{D}_{,i} \rangle \neq \langle \mathbf{D} \rangle_{,i}$. Then we have

$$2\mathcal{L}_{,i} = \text{Tr} [\mathbf{C}^{-1} \mathbf{C}_{,i} (\mathbf{I} - \mathbf{C}^{-1} \mathbf{D}) + \mathbf{C}^{-1} \mathbf{D}_{,i}], \quad (13.71)$$

($\mathbf{I}$ is the identity matrix) which averages to zero,

$$\langle \mathcal{L}_{,i} \rangle = 0. \quad (13.72)$$

This result is actually true for any distribution, not just Gaussian, since it corresponds to the derivative with respect to the parameters of the norm of the distribution. Notice that the average only acts on $\mathbf{D}$ since the random variables, the data, are only there, while of course derivatives act only on $\mathbf{C}$ and $\boldsymbol{\mu}$ since parameters are only there. To evaluate $\langle \mathcal{L}_{,ij} \rangle$ we notice that all first derivatives $\langle \mathbf{D}_{,i} \rangle$ vanish and that $\langle \mathbf{I} - \mathbf{C}^{-1} \mathbf{D} \rangle = 0$. Then we are finally left with [550, 551]

$$\begin{aligned} F_{ij} \equiv \langle \mathcal{L}_{,ij} \rangle &= \frac{1}{2} \text{Tr} \left[\mathbf{C}^{-1} \mathbf{C}_{,i} \mathbf{C}^{-1} \mathbf{C}_{,j} + \mathbf{C}^{-1} \langle \mathbf{D}_{,ij} \rangle \right] \\ &= \frac{1}{2} C_{\ell m}^{-1} \frac{\partial C_{mn}}{\partial \theta_i} C_{np}^{-1} \frac{\partial C_{p\ell}}{\partial \theta_j} + C_{\ell m}^{-1} \frac{\partial \mu_\ell}{\partial \theta_i} \frac{\partial \mu_m}{\partial \theta_j} \end{aligned} \quad (13.73)$$

(sum over repeated indices) where in the last equality we have written down the full index expression to be more explicit. Equation (13.63) is recovered when $\boldsymbol{\mu} = \mathbf{0}$ and $C_{\ell m} = \Delta_m^2 \delta_{\ell m}$.

13.5 Principal component analysis

So far we have almost always assumed a very specific model, for instance a model of the equation of state $w_{\text{DE}}(z)$, and have proceeded to get constraints on the parameters. The likelihood method will certainly find some constraints, no matter how wrong is our modeling. For instance, take the expansion $w_{\text{DE}}(z) = w_0 + w_1 z + w_2 z^2 + \dots$ and suppose that we stop at w_1 . Given a dataset of SN Ia, we could end up with very good constraints on w_0 but very loose on w_1 . We may content ourselves with that and blame the experimenters for their poor data. However, how can we be sure that the data do not contain good constraints on, say, w_2 or some other higher-order parameters? If the data do not extend very far we do not expect this, but still it would be nice to quantify which parameters (and how many) we can reasonably constrain for a given dataset. In other words we would like to *find* the best parametrization, rather than to *assume* one.

One way of doing this is to approximate the function $w_{\text{DE}}(z)$ in the range z_a, z_b with many stepwise constant values:

$$1 + w_{\text{DE}}(z) = \sum_{i=1}^N \theta_i(z) w_i, \quad (13.74)$$

where $\theta_i = 1$ for z inside the bin $(z_i, z_i + \Delta z)$ and 0 outside. So now we have $N (\gg 1)$ parameters w_i instead of two or three. Technically, this is just a bigger Fisher matrix problem and we could proceed as before. In this case, however, it would be

really nice to have uncorrelated errors on the parameters, since they all measure the same quantity, w_{DE} , and it will be difficult to compare different experiments if the errors are correlated (and compactifying to a single FOM would discard too much information). What we would like is in fact an expansion [552, 553]

$$1 + w_{\text{DE}}(z) = \sum_{i=1}^N \alpha_i e_i(z), \quad (13.75)$$

where the coefficients α_i are uncorrelated. Since uncorrelated parameters mean a diagonal Fisher matrix, the problem is solved by diagonalizing the Fisher matrix for the original N parameters w_i , thus obtaining a diagonal F_{ij}^D . The orthogonal basis functions $e_i(z)$ will be then the eigenvectors, with N eigenvalues λ_i . The new parameters α_i will have the variance $\sigma_i^2 = 1/\lambda_i = [(F^D)^{-1}]_{ii}$ (i.e. the elements on the diagonal of the inverse Fisher matrix).

Now, a parameter with a large error is a badly measured parameter. It means that the data are not able to measure that parameter very well. On the contrary, a parameter with small error is well measured. Therefore we can rank the parameters α_i according to their errors, that is, according to the magnitude of the eigenvalues of F_{ij} . The highest eigenvalues (smallest errors) are called “principal components” and the whole method is called *principal component analysis* (PCA). This method is based on the fact that every well-behaved function can be expanded in piecewise constant fragments and that every non-singular Fisher matrix can be diagonalized. That is, the PCA can always be used when we need to reconstruct an unknown function.

So we have now a few well-measured components plus many others with large errors. The eigenvectors $e_i(z)$ associated with the principal components are functions of z , built up by linear combinations of the $\theta_i(z_i)$. They tell us the range of z which is *best measured* by the data. We can plot them and have at once a view of the range of z most sensitive to that particular dataset, see Fig. 13.2. This is perhaps the best feature of the PCA since it allows us to optimize an experiment towards any range we are interested in.

The coefficients α_i themselves are rarely interesting. They can be evaluated by employing the property that the eigenvectors are orthogonal. Let us also normalize them by

$$\int e_i^2(z) dz = 1, \quad (13.76)$$

where the integration is taken in the whole z_a, z_b region. Multiplying Eq. (13.75) by $e_i(z)$ and then integrating, we obtain

$$\alpha_i = \int [1 + w_{\text{DE}}(z)] e_i(z) dz. \quad (13.77)$$

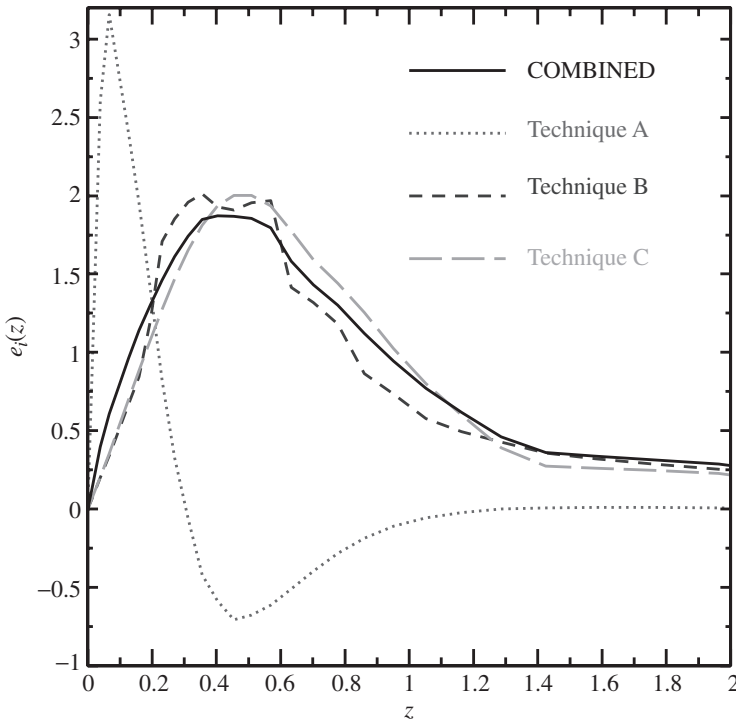


Figure 13.2 An example of the first principal component for three individual observational probes, and all of the techniques combined. From Ref. [554].

In comparing different experiments PCA might help, but care has to be taken when interpreting the results. In general the distribution of eigenvalues can be very different among different experiments and it is not obvious whether it is preferable to have a few well-measured components in a small range or many not-so-well-measured components in a large range. Reducing everything to a single FOM would kill the whole spirit of the PCA method and at the end of the day a sensible theoretical expectation is *the* principal component of any analysis.

13.6 Problems

- 13.1** Suppose 1% of the hypotheses are correct (in any sense one decides to define “correctness”). Suppose also that 80% of the times, an experiment testing a correct hypothesis confirms the hypothesis. On the other hand, 10% of the times an experiment tests a wrong hypothesis, it confirms that hypothesis due to experimental errors. What is the probability that a hypothesis that is confirmed by data is correct?
- 13.2** We denote the observational data as m_i , their errors σ_i , and the theoretical predictions as μ_i , where we assume that μ_i depend on a number of parameters $\theta_1, \theta_2, \dots$.

Marginalize the following likelihoods over the multiplicative nuisance parameter α :

$$L_1 = \exp \left[-\frac{1}{2} \sum_i \frac{(\alpha m_i - \mu_i)^2}{\sigma_i^2} \right], \quad (13.78)$$

$$L_2 = \exp \left[-\frac{1}{2} \sum_i \frac{(m_i - \alpha \mu_i)^2}{\sigma_i^2} \right]. \quad (13.79)$$

- 13.3** Find the Bayes' ratio for two models: model A predicting that a quantity $\theta = 0$ with no free parameters, and model B which assigns θ a Gaussian prior distribution with 0 mean and variance Σ . Assume that we perform a measurement of θ described by a normal likelihood of standard deviation σ , and with the maximum likelihood value lying λ standard deviations away from 0, i.e. $|\theta_{\max}/\sigma| = \lambda$. (From Ref. [544]).

Future observational constraints on the nature of dark energy

The observables used to obtain information about the global properties of the Universe are not many: distances, background radiation, source positions and velocities, galaxy shapes as an indicator of lensing shear, galaxy or cluster densities, all of them as functions of redshift. All these observables can in principle be employed to constrain the properties of dark energy. Five methods emerged so far as possibly the best tools for exploring the property of dark energy: SN Ia, CMB, LSS (including BAO), weak lensing and galaxy clusters. In some cases this selection was based on the actual current performance (e.g., the SN Ia method that gave birth to the whole dark energy concept); in others, on good promises for the next decade (e.g., weak lensing).

In Section 5.2 we already addressed the SN Ia method in some detail. In this chapter we discuss the other techniques and their prospects. We also explain the potential of alternative methods such as age tests, gamma ray bursts, strong lensing, and redshift drift.

14.1 Dark energy and the CMB

The physics of CMB is a wonderful playground for cosmologists. The initial conditions are set by inflation; the evolution of perturbations involves the delicate interplay of photons, baryons, neutrinos, dark matter, and dark energy, all coupled either directly or via the gravitational field. Finally, the observation and the analysis of the anisotropies themselves also involve very interesting physics, mathematics, and statistics. The outcome of this grand tour of physics is a powerful probe of the Universe from its very early stages down to the present time, providing us with fundamental (and in some case unique) constraints on many cosmological issues, from the spectrum of primordial scalar and tensor perturbations to the mass of neutrinos and the reionization epoch.

There are three aspects of CMB that deal most directly with dark energy: the acoustic peak position, the integrated Sachs-Wolfe (ISW) spectrum and the cross-correlation between ISW and LSS. We already discussed the peak position in Section 5.3. Here we discuss the ISW and ISW-LSS effects. A summary of WMAP results in the $(\Omega_m^{(0)}, \Omega_\Lambda^{(0)})$ plane is illustrated in Fig. 14.1.

14.1.1 Cross-correlation ISW-matter fluctuations

As we have seen in Section 4.11.1, the temperature anisotropy in the direction $\theta = (\theta_x, \theta_y)$ due to the ISW effect is an integral of the time variation of the quantity $\psi \equiv \Phi - \Psi$ along the line-of-sight:

$$\left. \frac{\delta T}{T} \right|_{\text{ISW}} \equiv \delta_T(\theta_x, \theta_y) = - \int_{\eta_i}^{\eta_0} d\eta e^{-\tau_{\text{op}}(\eta)} \frac{\partial \psi}{\partial \eta}, \quad (14.1)$$

where the conformal time η is integrated from some early pre-recombination epoch to the present. The only difference relative to the expression in Section 4.11.1 is the optical thickness factor $e^{-\tau_{\text{op}}(\eta)}$, where the optical depth $\tau_{\text{op}}(\eta)$ is defined in Eq. (4.168). The function $e^{-\tau_{\text{op}}(\eta)}$ weighs the photon path so that every $d\eta$ bin counts proportionally to its optical thickness.

In the context of photon propagation in linearly perturbed spaces we always consider photons propagating almost radially, i.e. for small deviation angles θ , and then we can put $x \approx r\theta \cos \phi \equiv r\theta_x$ and $y \approx r\theta \sin \phi \equiv r\theta_y$. Therefore one can think of the $dx^2 + dy^2$ part as the approximate Cartesian version of the angular interval $r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2$. In a non-flat metric as Eq. (2.54) we have instead $x = f_K(\chi)\theta_x$, $y = f_K(\chi)\theta_y$ where $f_K(\chi)$ is defined in Eq. (2.56) and χ is the comoving distance

$$\chi = \eta_0 - \eta, \quad (14.2)$$

which also corresponds to the look-back conformal time. Unless otherwise specified we work however in flat space and put $\chi = r$ so that $d\chi = dr = -d\eta$.

The ISW fluctuations due to the line-of-sight propagation of CMB photons are superimposed on the last-scattering surface fluctuations and are therefore difficult to disentangle. However, since the gravitational potentials Ψ and Φ are traced by the density fluctuations δ , we expect a non-zero correlation between δ_T and δ , as first proposed in Ref. [555]. The density fluctuations should include contributions from all sources, including dark energy fluctuations. The density fluctuations along the line-of-sight direction θ are

$$\delta(\theta_x, \theta_y) = \int dr W(r) \delta(\mathbf{r}), \quad (14.3)$$

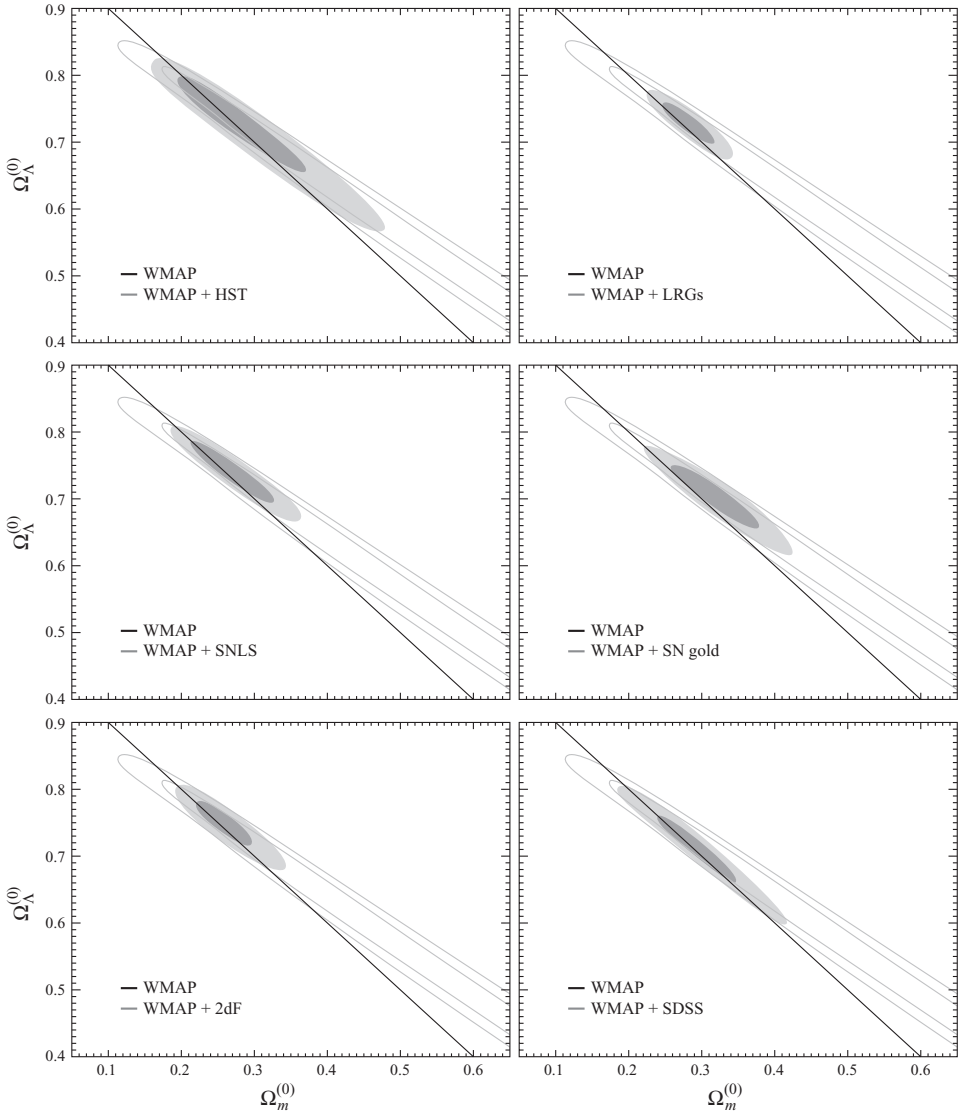


Figure 14.1 Marginalized contours (68% and 95%) for matter density $\Omega_m^{(0)}$ and vacuum energy density $\Omega_{\Lambda}^{(0)}$. The panels show various combinations of WMAP3yr and other data sets. (Upper left) WMAP + HST key project measurement of H_0 . (Upper right) WMAP + SDSS LRG measurement of the angular diameter distance to $z = 0.35$. (Middle left) WMAP + SNLS data. (Middle right) WMAP + SN Ia Gold data. (Lower left) WMAP + 2dF. (Lower right) WMAP + SDSS. From Ref. [14].

where as usual $W(r)$ is the selection function normalized so that $\int_0^\infty W(r)dr = 1$. All the r integrals in this subsection and the next one extend from 0 to the last scattering surface. If matter is represented by galaxies then $W(r)$ is the differential number density dN_g/dr of galaxies normalized to unity. In this case δ will be in fact δ_g/b .

Now we expand $\delta(\theta)$ in spherical harmonics with coefficients

$$\begin{aligned} a_{\ell m} &= \int d^2\theta Y_{\ell m}^*(\theta) \delta(\theta) \\ &= (2\pi)^{-3} \int dr d^2\theta d^3k Y_{\ell m}^*(\theta) W(r) e^{ik \cdot r} \delta(k) \\ &= (2\pi)^{-3} \int dr d^3k W(r) \delta(k) \int d^2\theta Y_{\ell m}^*(\theta) e^{ik \cdot r}, \end{aligned} \quad (14.4)$$

where we expanded the field in Fourier modes and assumed statistically isotropic fluctuations. As usual we discard the volume factors in the Fourier transforms. Now we can use the relativistic Poisson equation (4.64) and (4.59) to write

$$\psi = \Phi - \Psi = 3 \frac{\mathcal{H}^2}{k^2} \delta, \quad (14.5)$$

where δ is the total matter variable. This equation links the density fluctuations to the potential ones in standard gravity. If the only contribution to fluctuations comes from pressureless matter then $\delta = \Omega_m \delta_m$ and we can use the growth factor $D(\eta)$ normalized to present

$$\delta_m(\eta, k) = D(\eta) \delta_m^{(0)}(k), \quad (14.6)$$

to write the Fourier modes of $\partial\psi/\partial\eta$ as

$$\frac{\partial\psi(\eta, k)}{\partial\eta} = \frac{3(\mathcal{H}^2 \Omega_m D)_{,\eta}}{k^2} \delta_m^{(0)}(k). \quad (14.7)$$

The factor $(\mathcal{H}^2 \Omega_m D)_{,\eta}$ evaluates to zero for an Einstein–de Sitter Universe and consequently there is no cross-correlation. From Eqs. (14.1), (14.4), and (14.7) we can write the multipole coefficients for matter and ISW fluctuations in a unified and compact way:

$$a_{\ell n}^{(m,T)} = (2\pi)^{-3} \int dr d^3k W_{m,T}(r) \delta_m^{(0)}(k) \int d^2\theta Y_{\ell n}^*(\theta) e^{ik \cdot r}, \quad (14.8)$$

where

$$W_m \equiv W D, \quad W_T = 3k^{-2} (\mathcal{H}^2 \Omega_m D)_{,\eta}. \quad (14.9)$$

If the optical depth is not negligible, there is an extra factor of $e^{-\tau_{\text{op}}(\eta)}$. For modified gravity we have instead $W_T = 3k^{-2} (\Sigma \mathcal{H}^2 \Omega_m D)_{,\eta}$, where Σ has been defined in

(11.203). We need now a classical result for spherical harmonics:

$$\int e^{i\mathbf{k}\cdot\mathbf{r}} Y_{n\ell}^*(\hat{\mathbf{r}}) d^2\theta = 4\pi i^\ell j_\ell(kr) Y_{n\ell}^*(\hat{\mathbf{k}}), \quad (14.10)$$

where $\hat{\mathbf{r}} = \mathbf{r}/r$, $\hat{\mathbf{k}} = \mathbf{k}/k$ are unit vectors and $j_\ell(kr)$ are spherical Bessel functions. This can be obtained by first expanding the plane wave in spherical harmonics (see equation 16.127 in Jackson's textbook [556]):

$$e^{i\mathbf{k}\cdot\mathbf{r}} = 4\pi \sum_{\ell n} i^\ell j_\ell(kr) Y_{n\ell}^*(\hat{\mathbf{k}}) Y_{n\ell}(\hat{\mathbf{r}}), \quad (14.11)$$

and by integrating over the angles. Then we have

$$a_{\ell n}^{(m,T)} = \frac{i^\ell}{2\pi^2} \int d\mathbf{r} d^3k W_{m,T} j_\ell(kr) Y_{n\ell}^*(\hat{\mathbf{k}}) \delta_m^{(0)}(\mathbf{k}). \quad (14.12)$$

The multipole cross-correlation spectrum is then [555, 557, 558, 559, 560]

$$\begin{aligned} C_\ell &\equiv \langle a_{\ell n}^{(m)} a_{\ell n}^{(T)*} \rangle \\ &= \frac{2}{\pi} \int d\mathbf{r}_1 j_\ell(kr_1) W_m \int d\mathbf{r}_2 j_\ell(kr_2) W_T \int k^2 dk P_\delta^{(0)}(k) \int d^2\theta Y_{n\ell}^*(\hat{\mathbf{k}}) Y_{n\ell}(\hat{\mathbf{k}}), \end{aligned} \quad (14.13)$$

where we have used

$$\langle \delta_m(\mathbf{k}, \eta_0) \delta_m^*(\mathbf{k}', \eta_0) \rangle = (2\pi)^3 \delta_D^{(3)}(\mathbf{k} - \mathbf{k}') P_\delta^{(0)}(k). \quad (14.14)$$

Finally, the last integral in C_ℓ evaluates to unity due to the orthonormality of the spherical harmonics. Then we obtain

$$C_\ell = \frac{2}{\pi} \int_0^\infty k^2 dk I_\ell^{\text{ISW}} I_\ell^m P_\delta^{(0)}(k), \quad (14.15)$$

where

$$I_\ell^{\text{ISW}} = \int d\mathbf{r} j_\ell(kr) W_T = 3k^{-2} \int d\mathbf{r} e^{-\tau_{\text{op}}(\eta)} j_\ell(kr) (\Sigma \mathcal{H}^2 \Omega_m D)_{,\eta}, \quad (14.16)$$

$$I_\ell^m = \int d\mathbf{r} j_\ell(kr) W D. \quad (14.17)$$

If dark energy also contributes to the fluctuations then we can put $\Omega_t = 1$ in place of Ω_m and write

$$I_\ell^{\text{ISW}} = \int d\mathbf{r} j_\ell(kr) W_T = 3k^{-2} \int d\mathbf{r} e^{-\tau_{\text{op}}(\eta)} j_\ell(kr) (\Sigma \mathcal{H}^2 D_t)_{,\eta}, \quad (14.18)$$

$$I_\ell^t = \int d\mathbf{r} j_\ell(kr) W D_t, \quad (14.19)$$

where D_t is the growth factor of the total matter, $D_t = \delta(k, \eta)/\delta^{(0)}(k)$.

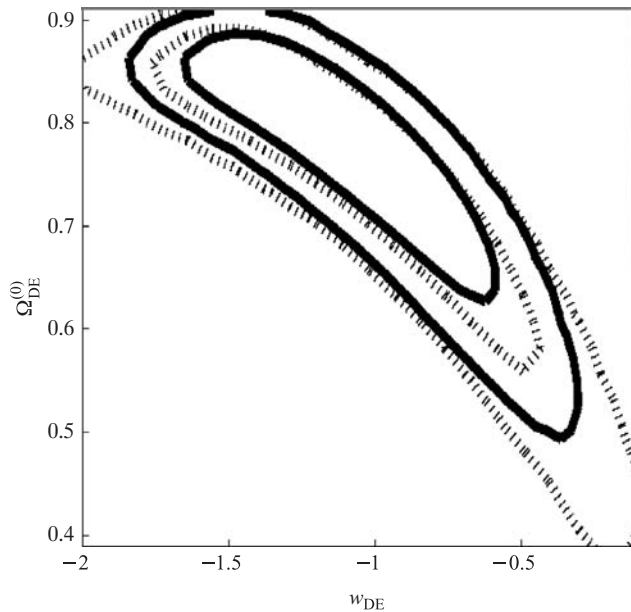


Figure 14.2 Marginalized contours in the $(w_{\text{DE}}, \Omega_{\text{DE}}^{(0)})$ plane for CDM plus dark energy with $w_{\text{DE}} = \text{constant}$ model (68% and 95% confidence level). The solid (dashed) line represents constraints from using WMAP+ISW+weak lensing (WMAP alone). From Ref. [562].

Note that Σ and D in general can be functions of both k and η . For completeness in I_{ℓ}^{ISW} we reintroduced the optical depth. The r integrals extend from $r = 0$ to some early epoch when the fluctuations were negligible. If the window functions $W_{m,T}$ are given in redshift, one has to change from r to z by using the relation $dr/dz = 1/H(z)$ in flat space.

It is now possible to use the cross-correlation ISW signal to test dark energy models. The theory gives us $P_{\delta}^{(0)}$, $\mathcal{H}(\eta)$, $\Omega_m(\eta)$, $D(k, \eta)$ as functions of model parameters (we also need $\tau_{\text{op}}(z)$ but for z less than a few we may put $\tau_{\text{op}} \approx 0$). The data are obtained by cross-correlating various large-scale surveys (e.g., the SDSS optical galaxies, the 2MASS infrared catalogue, the HEAO X-ray maps etc.) and the CMB sky maps. As usual, the galaxy fluctuations must be corrected for the bias before comparing to the predictions. In Fig. 14.2 we present the results obtained combining ISW with other probes. The signal has been detected to good significance ($4\text{--}5\sigma$), but the constraints on dark energy parameters are rather weak and strongly depend on the assumed sound speed of dark energy [561, 562]. Future large-scale surveys will no doubt improve considerably [563].

As a final note, we remark that the derivation of Eq. (14.15) is very general: it provides the link between the isotropic power spectrum $P(k)$ of a three-dimensional

field $\delta(x)$ weighted according to a unit-normalized radial selection function $W(r)$ and its multipole spectrum C_ℓ :

$$C_\ell = \frac{2}{\pi} \int_0^\infty k^2 dk P(k) \left[\int_0^\infty dr j_\ell(kr) W(r) \right]^2. \quad (14.20)$$

We can derive also another similar expression (that we already used in Section 5.3) in the following way. Let us assume that the window function $W(r)$ is a Dirac delta centered on the last scattering surface at distance $r = r^*$. It is possible to write the fluctuations as $\delta(\mathbf{k}) = \delta_i(\mathbf{k}) \hat{\delta}(k, \mu_{kr})$, where δ_i is the initial fluctuation. From Eq. (14.4) we then obtain

$$a_{\ell m} = (2\pi)^{-3} \int \delta_i(\mathbf{k}) d^3k \int d^2\theta \hat{\delta}(k, \mu_{kr}) Y_{\ell m}^*(\hat{\mathbf{r}}) e^{i\mathbf{k} \cdot \mathbf{r}^*}, \quad (14.21)$$

where $\hat{\delta}(k, \mu_{kr})$ can be developed in Legendre series:

$$\hat{\delta}(k, \mu_{kr}) = \sum_\ell (-i)^\ell (2\ell + 1) \mathcal{P}_\ell(\mu_{kr}) \delta_\ell(k). \quad (14.22)$$

Now we use the power spectrum definition $\langle \delta_i(\mathbf{k}) \delta_i^*(\mathbf{k}') \rangle = (2\pi)^3 \delta_D(\mathbf{k} - \mathbf{k}') P_i(k)$. Without loss of generality it is possible to set $P_i = 1$ since we can absorb in $\hat{\delta}$ the initial amplitude.

Squaring and averaging $a_{\ell m}$, we obtain

$$C_\ell = (2\pi)^{-3} \int d^3k \sum_{\ell\ell'} (-i)^\ell (i)^{\ell'} (2\ell + 1)(2\ell' + 1) \delta_\ell \delta_{\ell'}^* I(k) I^*(k), \quad (14.23)$$

where

$$I(k) = \int d^2\theta Y_{\ell m}^*(\hat{\mathbf{r}}) \mathcal{P}_\ell(\mu_{kr}) = \frac{4\pi}{2\ell + 1} Y_{\ell m}(\hat{\mathbf{k}}) \delta_{\ell\ell'}. \quad (14.24)$$

Finally, by integrating over the angles, we have

$$C_\ell = \frac{2}{\pi} \int_0^\infty dk k^2 |\delta_\ell|^2, \quad (14.25)$$

which corresponds to Eq. (5.28) with Θ_ℓ in place of δ_ℓ .

14.1.2 ISW effect on the CMB

In the previous section we have discussed the cross-correlation between galaxies and ISW. Here we consider instead the power spectrum of ISW alone as can be seen on the microwave background. From Eq. (14.15) it is clear that the result for

the correlation of the ISW with itself is

$$C_\ell = \frac{2}{\pi} \int k^2 dk P_\delta(k) (I_\ell^{\text{ISW}})^2. \quad (14.26)$$

An approximate result can be obtained by estimating

$$\begin{aligned} I_\ell^{\text{ISW}} &\approx \frac{3}{k^2} e^{-\tau_{\text{op}}(z_p)} (\Sigma \mathcal{H}^2 \Omega_m D)_{,\eta_p} \int dr j_\ell(kr) \\ &\approx \frac{3\sqrt{\pi}}{2k^3} e^{-\tau_{\text{op}}(z_p)} (\Sigma \mathcal{H}^2 \Omega_m D)_{,\eta_p} \frac{\Gamma[(\ell+1)/2]}{\Gamma[(\ell+2)/2]}, \end{aligned} \quad (14.27)$$

[see Eq. (17.11) in the Mathematical Appendix] where we have approximated the integral by taking $(0, \infty)$ as limits and where η_p is the conformal time evaluated at the peak of the spherical Bessel function, which can be obtained by the formula

$$r(z_p) = \eta_p - \eta_0 = \frac{\ell + 1/2}{k}, \quad (14.28)$$

so that η_p or $z_p = z(\eta_p)$ are functions of k and ℓ .

Then we have

$$C_\ell = \frac{9}{2} \left(\frac{\Gamma[(\ell+1)/2]}{\Gamma[(\ell+2)/2]} \right)^2 \int \frac{dk}{k^4} P_\delta(k) [e^{-\tau_{\text{op}}(z_p)} (\Sigma \mathcal{H}^2 \Omega_m D)_{,\eta_p}]^2. \quad (14.29)$$

The functions D and Σ have to be obtained by solving the general perturbation equations we have discussed in Section 11.2 (see problem 14.1). The ISW effect is in general important for small ℓ 's ($\ell \lesssim 10$) and it appears as a tilt of the low ℓ spectrum in the CMB (see Fig. 14.3). This contribution of the ISW to the CMB anisotropy spectrum is superimposed on all the others and it is not possible to single it out in general terms.

There are many effects that determine the amplitude of the ISW tail on the CMB spectrum. The potential ψ can vary in time due to either the background expansion $\mathcal{H}$ or the perturbation growth D . The two effects compensate exactly for a standard matter-dominated expansion but not in general. The total perturbation δ receives contributions also from dark energy perturbations, which are controlled by the dark energy sound speed c_s . In general, for the same background expansion and for $w_{\text{DE}} > -1$, if c_s^2 decreases, the dark energy clusters more and more like dark matter. This leads to the suppression of the ISW effect [564, 557].

On its own, the ISW effect is not a strong test of dark energy because it occurs at large angular scales, or low ℓ 's. Since we have a single sky, there are but a few “independent” large angle patches we can use to constrain the model. In other words, the low multipoles are affected by a large cosmic variance. Only rather extreme models can be convincingly ruled out by the ISW effect, e.g., models with long epochs of acceleration with fast growth of structure [486].

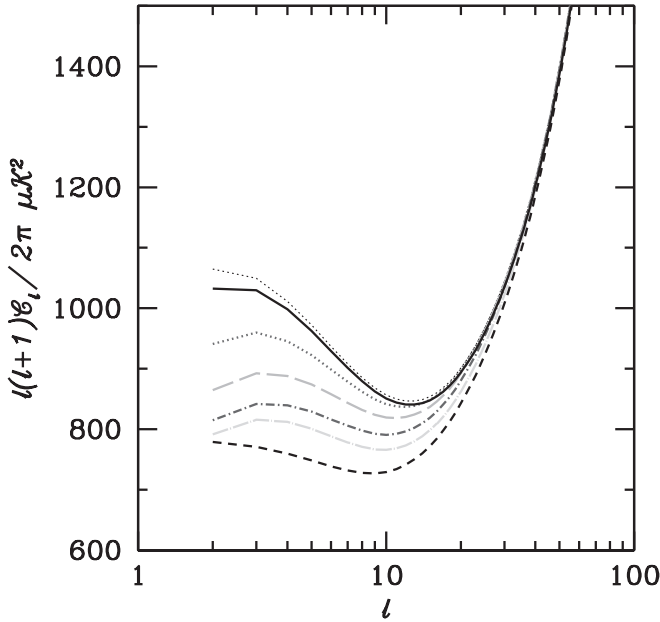


Figure 14.3 CMB anisotropies for CDM plus dark energy with $w_{\text{DE}} = -0.6$ with several different values of the sound speed. From top to bottom each line corresponds to $c_s^2 = 5, 1, 0.2, 0.05, 0.01, 0.0$ with dark energy perturbations, whereas the low dashed line corresponds to $c_s^2 = 1$ without dark energy perturbations. From Ref. [564].

14.2 Large-scale structure

We have seen in Section 5.4 that the baryons leave a non-negligible imprint on the total perturbations in the form of a coherent pattern of fluctuations at a fixed physical scale, see Fig. 14.4. This is only one example of physical mechanisms that can be observed in the matter power spectrum and that depend on cosmological parameters. Other features of the power spectrum that depend on cosmology are the position of the overall peak (or turnaround), the overall amplitude, and the slope. The problem we want to tackle in this section is how to extract all the cosmological information from the power spectrum. As before, we use the Fisher matrix approach. Most of the discussion in this section is based on the work of Seo and Eisenstein [549, 565].

One obvious way would be to calculate from scratch the power spectrum for all values of the cosmological parameters we want to study. However this would obscure the dependence of the parameters and could require long computations. Here we illustrate a way to make the dependence on cosmology as explicit as possible. The problem at hand is the following. If we know the power spectrum at redshift $z = 0$ for a given reference or fiducial cosmology (subscript r), e.g., the

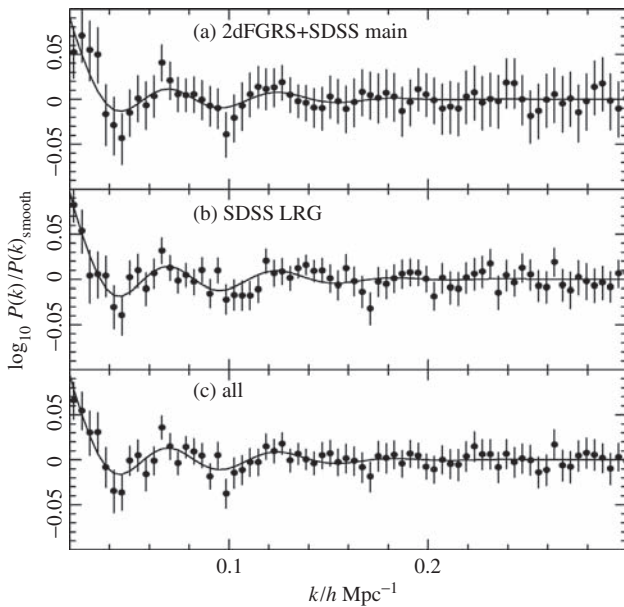


Figure 14.4 Ratio of the full observed power spectra to suitably smoothed spectra to show the oscillating BAO component. Power spectra are calculated from (a) the combined SDSS and 2dFGRS main galaxies, (b) the SDSS DR5 LRG sample, and (c) the combination of these two samples (solid symbols with 1σ errors). The solid line in each panel is a fit to the data. From Ref. [124].

standard Λ CDM, how do we find the spectrum for any other cosmology at any other redshift?

The cosmological model influences the spectrum in many ways. It changes the shape of the spectrum at $z = 0$. It changes the amplitude of the spectrum at any z through the growth factor. It affects the separation between galaxies and therefore also the wavenumbers k in the spectrum. Finally, it also changes the volume in which the spectrum is calculated. Let us discuss the latter two effects first.

If we observe an angle θ subtending a transverse comoving scale λ_1 at z , then the angular diameter distance is $d_1(z) = \lambda_1/(1+z)\theta$ where the subscript 1 indicates a given cosmology, i.e. some values of $\Omega_m^{(0)}$, $\Omega_\Lambda^{(0)}$ etc. In a different cosmology (subscript 2), the relation will be $d_2(z) = \lambda_2/(1+z)\theta$, i.e. the scale has to change in order to keep the same subtending angle at the same redshift. It follows that for any cosmology the combination d/λ for each given angle is a constant. The same goes for the combination $k_\perp d$ if $k_\perp$ is the transverse wavenumber corresponding to that transverse scale. Therefore, if we take a reference cosmology r , we have that for any other cosmology the *transverse* wavenumber is given by

$$k_\perp = k_{r\perp} d_r / d. \quad (14.30)$$

A similar argument can be applied to the comoving scale extending *along* the line of sight from z_1 to z_2 . The scale is then $\lambda = dz/H(z)$ and in order for this scale to be seen at the same $dz = z_2 - z_1$ the product λH has to remain constant when changing cosmology. Therefore, along with (14.30), we have for radial modes

$$k_{\parallel} = k_{r\parallel} H/H_r. \quad (14.31)$$

Clearly, any wave vector $\mathbf{k}$ can be decomposed into $k_{\parallel}$ and $k_{\perp}$. The relations above apply therefore to any perturbation mode. Every mode $\mathbf{k}$ in the power spectrum can be written in terms of the reference mode $\mathbf{k}_r$ with an explicit dependence on the cosmological parameters inside d and H . We know then how the wavenumber changes with cosmology. This implies that if a power spectrum is isotropic for the reference cosmology, it will become anisotropic for any other cosmology, because $k_{\parallel}$ and $k_{\perp}$ change differently: this is called the Alcock–Paczynski effect [566]. However we are measuring more than just the anisotropy since if we have a feature in $P(k, z)$ at a *fixed* comoving scale then we have $\theta_1 d_A(z_1)(1 + z_1) = \theta_2 d_A(z_2)(1 + z_2)$ and we can measure the ratio

$$\frac{\theta_1}{\theta_2} = \frac{d_A(z_2)(1 + z_2)}{d_A(z_1)(1 + z_1)}, \quad (14.32)$$

and analogously the ratio $H(z_2)/H(z_1)$. And we do have such a feature: the baryon acoustic oscillations we have already encountered in Section 5.4 that manifest themselves as small oscillations superimposed on the smooth matter power spectrum. Therefore, by observing the power spectrum at two or more redshifts we can measure the wavevectors at which the baryon acoustic scale is located and determine separately the ratios of $H(z)$ and of $d_A(z)$. In the method we illustrate below we do not need to identify “by hand” the features we want to measure: all the scale-dependent information contained in the power spectrum will be automatically employed to constrain cosmology. To the galaxy clustering we can finally add the measure of the baryon acoustic oscillations on the CMB.

From the relations (14.30) and (14.31) we derive the relation between the wavenumber modulus k and the direction cosine $\mu = \mathbf{k} \cdot \mathbf{r}/k$ (here $\mathbf{r}$ is the unit vector parallel to the line of sight) in the reference cosmology and in the generic cosmology

$$k = (k_{\parallel}^2 + k_{\perp}^2)^{1/2} = R k_r, \quad (14.33)$$

$$\mu = \frac{k_{\parallel}}{(k_{\parallel}^2 + k_{\perp}^2)^{1/2}} = \frac{H \mu_r}{H_r R}, \quad (14.34)$$

where

$$R = \frac{\sqrt{H^2 d^2 \mu_r^2 - H_r^2 d_r^2 (\mu_r^2 - 1)}}{H_r d}. \quad (14.35)$$

Since the power spectrum is proportional to the volume V in which we measure the perturbations, we need to evaluate also how V depends on cosmology. If we measure the spectrum in a solid angle $\theta^2 \text{ rad}^2$ and a shell of thickness dz , then the volume is

$$V = \theta^2 r^2 dr = d^2(z) r_{,z}(z) dz = \frac{d^2}{H} dz, \quad (14.36)$$

where $r_{,z} = dr/dz = 1/H(z)$ and $d = r(z)\theta$. It follows that VH/d^2 is independent of the cosmology and therefore

$$V = V_r \frac{H_r d^2}{H d_r^2}. \quad (14.37)$$

For the purpose of obtaining a power spectrum from real data, we need to assume a reference cosmology to convert angles and redshifts into distances or wave vectors. The relations (14.33, 14.34, 14.37) above allow us to relate this reference power spectrum to a general power spectrum for any d , H , i.e. for any given cosmology. The power spectrum $P(k) = V \delta_k^2$ for the true cosmology can be converted into the power spectrum in the reference cosmology (the one we use to convert observed redshifts into distances) by multiplying by V_r/V and by converting k , μ into k_r , μ_r . Hence we can write at any redshift [567]

$$P_r(k_r, z) = \frac{H(z) d_r^2(z)}{H_r(z) d^2(z)} P(R k_r, z). \quad (14.38)$$

Notice that R depends on z .

As long as the perturbations grow in time independently of their wavelengths (which is the case in most simple scenarios), we can write the spectrum at any z by multiplying the present spectrum by the growth factor squared: $P(k, z) = D(z)^2 P(k, 0)$, where $D(z) = \delta_m(z)/\delta_m(0)$. Then, we can relate the observed galaxy power spectrum to the theoretical matter power spectrum by the bias factor $b^2(k, z)$. Finally, we must connect the observations in redshift space to the theoretical predictions which are performed in real space. As we have seen in Section 4.8 this requires an extra factor $(1 + \beta\mu^2)^2$. Putting everything together, we finally obtain

$$P_{r,\text{obs}}(k_r, \mu_r; z) = P_s(z) + \frac{H(z) d_r^2(z)}{H_r(z) d^2(z)} D^2(z) b^2(z) [1 + \beta(z) \mu^2]^2 P(k, z = 0). \quad (14.39)$$

We have also added $P_s(z)$, a scale-independent offset which can arise if our removal of the shot-noise [see Eq. (3.38)] is incomplete. It is always a good idea to imagine

possible sources of systematic error and to insert them in the Fisher analysis. Notice that we have assumed that b and β do not depend on the scale but only on z . However this assumption is not completely tested and is certainly false on small non-linear scales. The last factor on the r.h.s. of Eq. (14.39) is the matter power spectrum at $z = 0$, which in turn depends on the cosmological parameters ($\Omega_m^{(0)}$, $\Omega_b^{(0)}$, n_s , h , dark energy parameters, etc). These parameters are not as explicit as the others inside d , H and when we calculate in the Fisher matrix (13.64) the terms $\partial P_{r,\text{obs}}/\partial \Omega_m^{(0)}$, for instance, we have to remember to differentiate also $P(k, 0)$.

The parameter β is defined by $\beta = f/b$, where $f = \dot{\delta}_m/(H\delta_m)$ is the growth rate of matter perturbations. Whenever f is approximated to be $f \approx \Omega_m^\gamma(z)$, we can use γ as an extra free parameter (or fix it to 0.55 if we restrict ourselves to the Λ CDM model). If we assume this growth factor, then $D(z)$ evaluates to

$$D(z) = \exp \left[\int_z^0 \Omega_m^\gamma(\tilde{z}) \frac{d\tilde{z}}{1+\tilde{z}} \right]. \quad (14.40)$$

Note that sometimes in literature one meets with the different definition $D(a) \sim \delta/a$, instead of our $D \sim \delta$.

Finally, we wish to include a redshift error in the observed power spectrum. The typical redshift error for spectroscopic surveys is very small (less than 0.1%) and probably negligible, but in other cases (photometric or broad-band surveys) the redshift determination can be quite noisy, up to a few percent. To derive this extra source of error one can proceed as follows. Since $dr = dz/H(z)$, an error of σ_z in redshift measurement means an error of $\sigma_r = \sigma_z/H(z)$ in distance. Suppose that the observed radial distances r are Gaussian distributed around the true distances r_0 :

$$f(r, r_0) = \frac{1}{\sqrt{2\pi}\sigma_r} e^{-(r-r_0)^2/(2\sigma_r^2)}. \quad (14.41)$$

Then the observed correlation function is given by the convolution

$$\xi(\sigma, r_0) = \int_0^\infty \xi[\sigma, r] f(r, r_0) dr. \quad (14.42)$$

After the Fourier transformation, the convolution becomes a product:

$$P = P_{r,\text{obs}} e^{-k^2 \mu^2 \sigma_r^2}, \quad (14.43)$$

where

$$\sigma_r = \sigma_z/H(z). \quad (14.44)$$

The observed power spectrum depends therefore on a number of parameters, denoted collectively p_i , such as $\Omega_m^{(0)}$, H_0 , n_s , w_0 . Then we calculate, numerically

or analytically, the derivatives

$$\left(\frac{\partial \ln P}{\partial p_i} \right)_r, \quad (14.45)$$

evaluated for the reference (or “fiducial”) model. Finally, as we have already seen in Section 13.4 the Fisher matrix is

$$F_{ij} = \frac{1}{8\pi^2} \int_{-1}^1 d\mu \int_{k_{\min}}^{k_{\max}} k^2 dk \left(\frac{\partial \ln P}{\partial p_i} \frac{\partial \ln P}{\partial p_j} \right)_r V_{\text{eff}}(k, \mu), \quad (14.46)$$

where

$$V_{\text{eff}}(k, \mu) = \left[\frac{n P(k, \mu)}{n P(k, \mu) + 1} \right]^2 V_{\text{survey}}. \quad (14.47)$$

Note that n is the number density of galaxies. The small scale cut-off $k_{\max}$ should be such as to discard the non-linear part of the spectrum, since the theoretical prediction of the galaxy spectrum is rather uncertain due to the possible scale-dependence of the bias and hydrodynamical effects difficult to model properly with N -body simulations. Often this cut-off is realized by imposing $k_{\max} = \pi/2R_{\max}$ such that the amplitude $\sigma(R_{\max})$ defined in Eq. (3.58) is much smaller than unity, say, 0.3. The other limit $k_{\min}$ is instead less important in general and can be taken to vanish because the integrand goes rapidly to zero for small k .

As we have noticed, we can measure $P(k, z)$ for several redshift bins, i.e. grouping galaxies according to their redshift. In this case we have an independent Fisher matrix for each bin centered at z_n and the final Fisher matrix is their sum:

$$F_{ij}^{(\text{tot})} = \sum_n F_{ij}(z_n). \quad (14.48)$$

Clearly each bin will have its own expected density $n(z)$ and bias $b(z)$ that depend on the survey specification. The cosmological parameters p_i are typically $\Omega_m^{(0)}$, $\Omega_b^{(0)}$, h , n_s plus the dark energy parameters according to the model one is testing. One can also work with an intermediate set of parameters formed by (the logarithm is just for convenience)

$$d_i \equiv \ln d(z_i), \quad h_i \equiv \ln H(z_i), \quad g_i \equiv \ln[D(z_i)b(z_i)], \quad \beta_i \equiv \ln \beta(z_i), \quad (14.49)$$

so that we have four parameters for each redshift bin (to which we can add a shot noise parameter $P_{s,i}$ for each bin), plus the cosmological parameters that enter directly into the present spectrum $P(k, z=0)$. Of course here we are assuming that all these functions can be approximated as constant inside the bins, which therefore should be taken small enough for this to be approximately true, and yet populated enough for the noise not to dominate the signal. Then from this set one can project onto the final set of cosmological parameters following the steps in Section 13.4. Notice that D, b occur only as a product and therefore they form collectively a

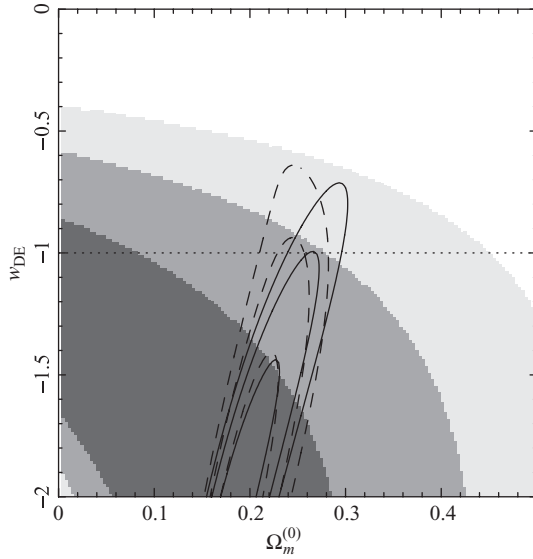


Figure 14.5 Likelihood surfaces in the $(\Omega_m^{(0)}, w_{DE})$ plane assuming a flat space with constant dark energy equation of state, for spectra in the 2dF and SDSS redshift surveys. The shaded regions show the likelihood from the redshift surveys alone. The solid contours are calculated by modeling the CMB sound horizon scale and the dashed contours by including the CMB peak position measurement. Notice how strong is the effect on the confidence regions on $\Omega_m^{(0)}$, but much weaker on w_{DE} . The dotted line shows $w_{DE} = -1$. From Ref. [124].

single parameter, $D \cdot b$, for each redshift bin. If we treat them separately then the Fisher matrix would be singular, since $\partial \ln P / \partial \ln D = \partial \ln P / \partial \ln b$ and therefore the D -row in F_{ij} would be equal to the b -row. In principle this degeneracy could be broken since b appears also in $\beta = f/b$. In general, however, since we know so little of the bias (for instance how it depends on scale and time) it appears safer to marginalize over it.

The ability to constrain the dark energy parameters depends crucially on combining several redshift bins. If we have a single bin, i.e. a single pair H_i, d_i , we see that the dark energy equation of state $w_{DE}(z_i)$ enters in an identical way in both so that there is full degeneracy between e.g., w_0, w_1 in the parametrization of $w_{DE}(z_i) = w_0 + w_1 z_i / (1 + z_i)$. The acoustic peaks in the CMB give us a $z \approx 1100$ bin to which we can compare present-day observations but the $w_{DE}(z)$ information at that redshift is very diluted so that we still need intermediate redshift spectra in order to exploit the full cosmological power of $P(k)$. Figure 14.5 shows that the effect of adding CMB information is important to reduce the overall errors, but the final uncertainty on w_{DE} is only weakly affected.

The Fisher method presented here can be easily adapted to include several other parameters, from a general growth factor to modified gravity parameters [569], or it can be made more robust by discarding part of the information contained in the power spectrum that is judged more prone to depend on uncontrolled effects. For instance the spectrum broad shape depends in general on other phenomena we have not included so far, like massive neutrinos or other light components, on the primordial initial slope, on non-linear corrections, and on the scale-dependence of the bias factor. So one may wish to discard this information and retain only the peak position, which being a pure geometrical effect is hard to spoil without invoking ad hoc phenomena. This can be done in several ways, all amounting to methods for extracting the oscillations from the spectrum, for instance smoothing the spectrum or subtracting the “continuum” [570, 571, 565]. Non-linear broadening of the wiggles could also be accounted for in a parametric way [565].

Next decade surveys plan to catalog tens or hundreds of million redshifts, so as to measure $P(k)$ at high redshift, up to $z \approx 2$. This will allow a reconstruction of H, d in bins and to unleash the power of large-scale structure to characterize dark energy. Using the specifications for a full-sky redshift survey of a total of 10^8 galaxies in the redshift range $z < 2$, it will be possible to find constraints on w_{DE} of the order of 0.01 and 0.1 respectively [568]. One example of the forecast is given in Fig. 14.6, which shows the observational constraints for future redshift surveys on the dark energy equation of state.

14.3 Growth function

We have seen that an ingredient of the observed power spectrum is the redshift distortion induced by the peculiar motion, embodied in the correction $1 + \beta\mu^2$. Although this term can be analyzed just as all the other ones in the full $P(k)$, it is also possible to isolate it from the other effects. The interest in doing so is that it can directly give us $\beta = f/b$ and, if we know the bias, the growth rate f . As we have seen, any deviation of $f = \Omega_m^\gamma$ from the standard value $\gamma \approx 0.55$ is a signal that something well beyond a simple cosmological constant or a slowly varying dark energy component is at work.

As we have studied in Section 4.8, this redshift correction generates an anisotropy in the correlation function or the power spectrum. Since we expect the galaxy distribution to be statistically isotropic, any deviation from an isotropic spectrum or correlation function can be used to estimate β . Most works in this area have been performed using directly the correlation function, rather than the spectrum [572, 573, 574], so we discuss this approach here.

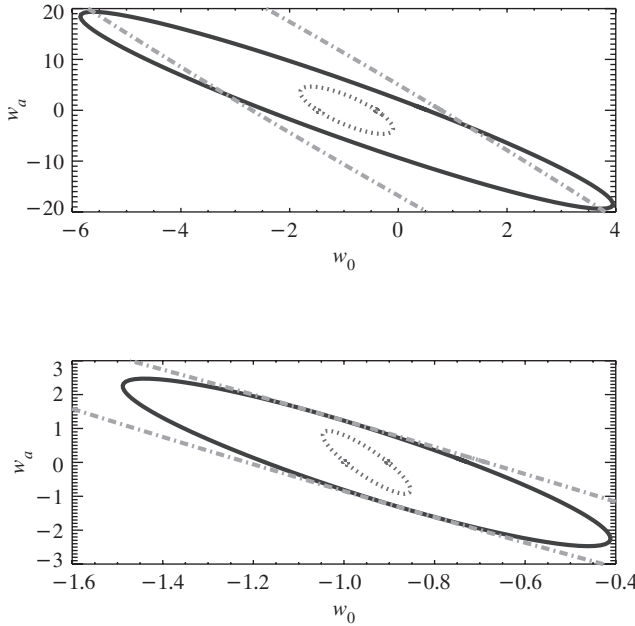


Figure 14.6 Examples of forecast of marginalized 1σ constraints on dark energy equation of state parameters w_0, w_a with the parametrization $w_{\text{DE}}(z) = w_0 + w_a z/(1+z)$ for future redshift surveys. Top: redshift survey from typical next-decade ground observations. Bottom: redshift survey from future space observations. The inner contour is for the full $P(k)$ method; the outer contours for methods that use less information. From Ref. [568].

We have seen in Section 4.8 that the relation between the power spectrum in real space (subscript r) and the spectrum in redshift space (subscript s) is given by

$$P_s(k) = P_r(k)(1 + \beta\mu^2)^2. \quad (14.50)$$

If k_π is the component of the wavenumber along the line-of-sight vector $\boldsymbol{\pi}$, then $\mu^2 = k_\pi^2/k^2$. We can derive a similar relation for the correlation simply by Fourier anti-transforming, i.e.

$$\xi_s(\mathbf{r}) = \int P_s(k) e^{i\mathbf{k}\cdot\mathbf{r}} d^3k = \int P_r(k) (1 + \beta\mu^2)^2 e^{i\mathbf{k}\cdot\mathbf{r}} d^3k. \quad (14.51)$$

By expanding $\xi_s(\mathbf{r})$ in Legendre polynomials $\mathcal{P}_\ell(\mu)$, you can show in problem 14.2 that [572]

$$\xi_s(r, \mu) = \xi_0(r)\mathcal{P}_0(\mu) + \xi_2(r)\mathcal{P}_2(\mu) + \xi_4(r)\mathcal{P}_4(\mu), \quad (14.52)$$

where μ is now the cosine angle between the line of sight and the separation $\mathbf{r}$, and

$$\begin{aligned}\xi_0 &= \left(1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2\right) \xi_r(r), & \xi_2 &= \left(\frac{4}{3}\beta + \frac{4}{7}\beta^2\right) [\xi_r(r) - I_2], \\ \xi_4 &= \frac{8}{35}\beta^2 \left[\xi_r(r) + \frac{5}{2}I_2 - \frac{7}{2}I_4\right],\end{aligned}\quad (14.53)$$

with

$$I_2 = 3r^{-3} \int_0^r \xi_r(y) y^2 dy, \quad I_4 = 5r^{-5} \int_0^r \xi_r(y) y^4 dy. \quad (14.54)$$

The first few Legendre polynomials are $\mathcal{P}_0 = 1$, $\mathcal{P}_2 = (3\mu^2 - 1)/2$, and $\mathcal{P}_4 = (35\mu^4 - 30\mu^2 + 3)/8$. We see that if $\beta = 0$ then ξ_s does not depend on μ and there is no distortion of the correlation function; any anisotropy of ξ_s can then be used to estimate β .

However there is an extra complication. The linear approach fails on small scales. In redshift galaxy catalogs we see clearly that in the core of clusters the galaxies appear distributed in an elongated way along the line of sight, due to the extra dispersion in redshift induced by the strong and random peculiar motions. This is the so-called “fingers-of-god” effect. Of course in principle we could simply limit ourselves to large scales, but the separation between large and small scales is not easy. On the other hand, we can estimate the velocity distribution in the high-density cores of clusters and use this information to derive the redshift distortion on small scales. It turns out observationally that a good fit to the galaxy velocity distribution at small scales is given by an exponential form

$$f(v) = \frac{1}{\sqrt{2}\sigma_v} e^{-\sqrt{2}|v|/\sigma_v}, \quad (14.55)$$

where $\sigma_v \approx 500$ km/sec. That is, we can assume that the observed line-of-sight separation π equals the true distance π_t plus an “error” due to the peculiar velocity v [see Eq. (4.31)]:

$$\pi = \pi_t + \frac{v(1+z)}{H(z)}, \quad (14.56)$$

distributed according to $f(v)$ (while the transverse component σ is unaffected). Therefore the observed correlation function is given by taking the theoretical one $\xi_s(\sigma, \pi_t)$ and convolving with the velocity distribution:

$$\xi_s(\sigma, \pi) = \int_0^\infty \xi_s \left[\sigma, \pi - \frac{v(1+z)}{H(z)} \right] f(v) dv. \quad (14.57)$$

In this way one obtains (numerically, unless one approximates ξ_s as a power law, which is rather doubtful) a correlation function $\xi_s(\sigma, \pi; \beta, \sigma_v)$ that depends

simultaneously on β , σ_v and that can be fitted to the real data. In Eq. (4.139) we found an empirical form for the Fourier-space version of (14.57).

But what do we assume as undistorted $\xi_r(r)$ in Eq. (14.52)? The linear matter power spectrum can be Fourier anti-transformed, but since we want to go beyond linearity this approach is not helpful. The simplest way to go is then not to assume but to estimate the undistorted correlation. And we have one: it is the transverse correlation, i.e. the correlation along σ . We can derive it by integrating away the π part:

$$\Xi(\sigma) = 2 \int_0^\infty \xi(\sigma, \pi) d\pi, \quad (14.58)$$

(the integrand can be equivalently in real or redshift space) where the factor of 2 appears because ξ is even in π . Setting $r^2 = \sigma^2 + \pi^2$, Eq. (14.58) can be written as

$$\Xi(\sigma) = 2 \int_\sigma^\infty \frac{r \xi_r(r)}{(r^2 - \sigma^2)^{1/2}} dr. \quad (14.59)$$

Finally, this equation can be inverted to give [576]

$$\xi_r(r) = -\frac{1}{\pi} \int_r^\infty \frac{(d\Xi(\sigma)/d\sigma)}{(\sigma^2 - r^2)^{1/2}} d\sigma, \quad (14.60)$$

which is called the Abel transform (you can verify it in problem 14.3). Hence we can derive from the data themselves the undistorted $\xi_r(r)$, plug it into Eq. (14.52), then perform the convolution given by (14.57) and finally fit for the parameters β and σ_v . An example of a correlation in redshift space that shows both the linear flattening and the fingers-of-god effect is illustrated in Fig. 14.7.

Naturally, all this sweeps under the rug a host of important assumptions: just to name a few, that σ is independent of scale and redshift, that β is independent of scale and redshift, and that the function $f(v)$ is also universal, not to speak of the problem of performing the inversion (14.60) which certainly cannot be carried out to infinity and which requires a noisy derivative of Ξ .

Presumably the most dangerous assumption we have made is that we know how to derive the separation between two sources from the observed redshifts. The distance to each source in general requires an integral over the geometry, which in flat space is given by $r(z) = \int_0^z d\tilde{z}/H(\tilde{z})$. This in turn requires the knowledge of cosmological parameters. If we assume a cosmological model from the beginning, our final result for β is bound to depend on this assumption. Any change in cosmological parameters will induce a change in r and, what is worse, a change that will affect the transverse and the radial separation in a different way. This is of course the same Alcock–Paczynski effect we have seen in the previous section in k -space. In fact, even beside the redshift distortion, the assumption of a wrong

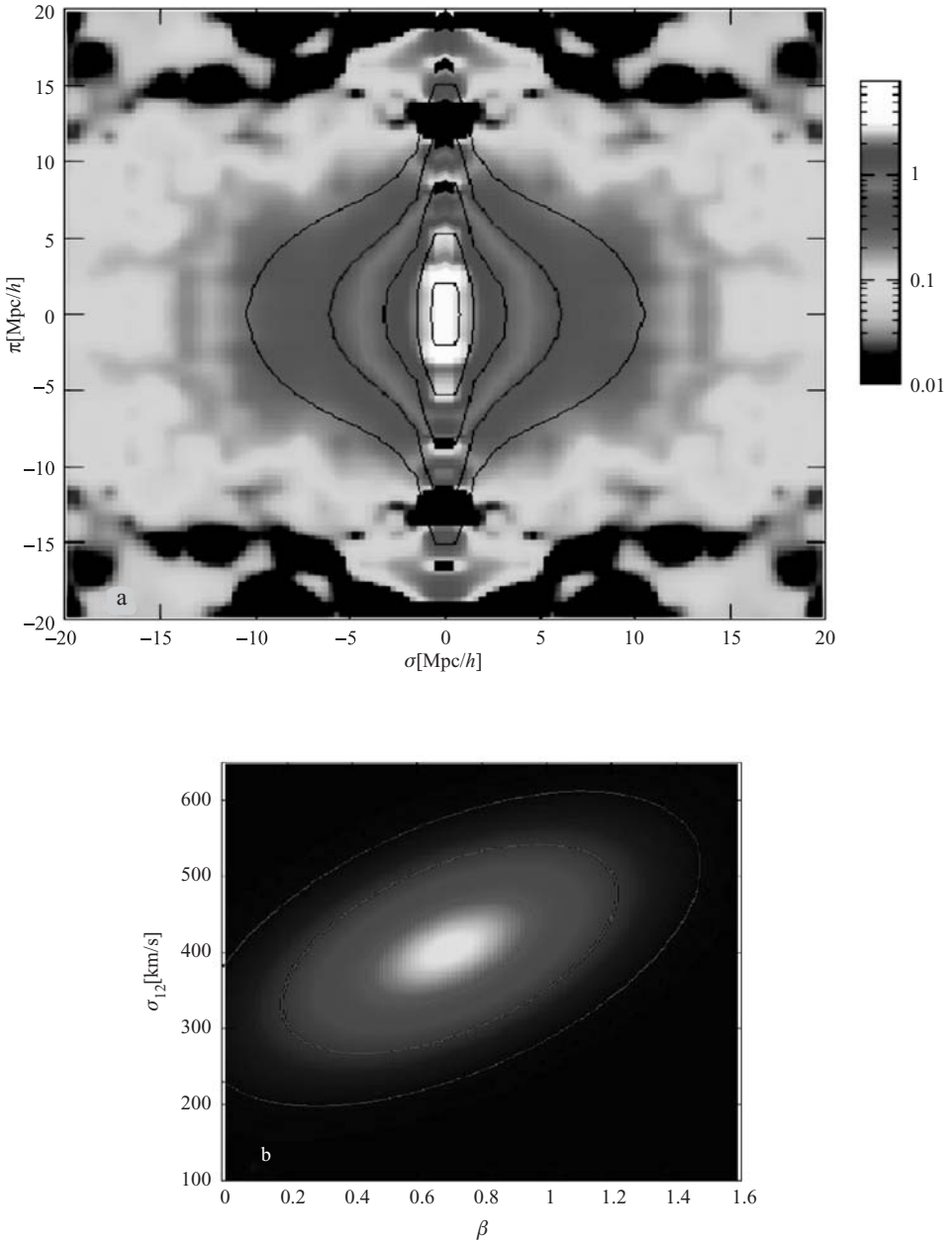


Figure 14.7 (Top panel) The correlation function $\xi(\sigma, \pi)$ as measured using approximately 6000 galaxy redshifts with $0.6 < z < 1.2$ (effective redshift 0.77) in the VVDS-Wide survey (replicated over four quadrants to show the deviations from circular symmetry). The superimposed solid contours correspond to the best-fitting distortion model with a distortion parameter $\beta = 0.7$ and a pairwise dispersion $\sigma_{12} = 412$ km/sec, obtained by maximizing the model likelihood given the data. (Bottom panel) Confidence levels (1,2,3 sigma) for the β and σ_{12} . (From Ref. [574]. Reprinted by permission from Macmillan Publishers Ltd.)

cosmological model by itself already induces an anisotropy on ξ so that one would be better off by including directly the cosmological parameters into the function $\xi(\sigma, \pi)$ or to revert to the full $P(k)$ method of the previous section.

14.4 Cosmic shear

The second pillar of future dark energy observations is likely to be the detection of cosmic shear induced by weak lensing (for very useful reviews see Refs. [577, 578]). This has already been observed in limited areas of the sky and there are plans to cover almost all-sky cosmic shear surveys (see Refs. [579, 580, 581, 582] for the first weak lensing detection and Ref. [583] for a review of recent observations). The great advantage of weak lensing is that it is due to the overall mass distribution and not to the luminous matter component alone. Contrary to the clustering method, it is therefore not affected by the biasing uncertainty.

Let us start again from the usual flat perturbed conformal metric

$$ds^2 = a^2(\eta)[-(1 + 2\Psi)d\eta^2 + (1 + 2\Phi)(dr^2 + dx^2 + dy^2)], \quad (14.61)$$

and consider a light ray propagating, in the unperturbed metric, along the r axis. The unusual coordinate notation r, x, y instead of x, y, z is useful to avoid confusion between the redshift and the coordinate z . For small angles θ we have as usual $x \approx r\theta \cos \phi$ and $y \approx r\theta \sin \phi$. This problem of photon propagation has been already addressed in Section 4.11.2. The conclusion was that the distortion of an image which is described in the *source* plane by the vector components $\theta_i^s = \{\theta_x, \theta_y\} = \{\theta^s \cos \phi^s, \theta^s \sin \phi^s\}$ and in the *observed* or lens plane by the analogous vector θ_i is entirely described at first-order by the transformation matrix A_{ij} given in Eq. (4.244).

Let us see how to calculate the distortion observationally. In practice, one measures the ellipticity of each galaxy in the survey, finds the correlation of the measures, and subtracts the expected random noise. A galaxy, for an observer, is just a function $I(\theta_x, \theta_y)$ of luminous intensity distributed over the sky, centered at $\theta = 0$. The quadrupole moment of the image is given by

$$q_{ij} = \int d^2\theta I(\theta) \theta_i \theta_j, \quad (14.62)$$

and two simple measures of ellipticity are

$$\varepsilon_1 = \frac{q_{xx} - q_{yy}}{q_{xx} + q_{yy}}, \quad \varepsilon_2 = \frac{2q_{xy}}{q_{xx} + q_{yy}}. \quad (14.63)$$

For small angles we have $\theta_i = (A^{-1})_{ij}\theta_j^s$, where [see Eq. (4.244)]

$$A_{ij} = \delta_{ij} + D_{ij} = \begin{pmatrix} 1 - \kappa_{\text{wl}} - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa_{\text{wl}} + \gamma_1 \end{pmatrix}, \quad (14.64)$$

and the ellipticity $\varepsilon_{1,2}$ can be expressed as a function of the distortion tensor A_{ij} and therefore of κ_{wl} , γ_1 , γ_2 . An intrinsically circular object, $I(\theta_x^s, \theta_y^s) = f[(\theta_x^s)^2 + (\theta_y^s)^2]$ (where $f(x)$ is an arbitrary function describing the luminous intensity), is distorted into an elliptical one whose luminosity distribution can be expanded at first-order for small κ_{wl} , $\gamma_{1,2}$:

$$\begin{aligned} I(\theta_x, \theta_y) &= f[(A_{xj}\theta_j)^2 + (A_{yj}\theta_j)^2] \\ &\approx f(\theta^2) + 2\kappa_{\text{wl}}\theta^2 f_{,\theta^2}(\theta^2) - 2(\gamma_1\theta_x^2 - \gamma_1\theta_y^2 + 2\gamma_2\theta_x\theta_y)f_{,\theta^2}(\theta^2), \end{aligned} \quad (14.65)$$

where $\theta^2 = \theta_x^2 + \theta_y^2$ and $f_{,\theta^2} = \text{d}f/\text{d}(\theta^2)$. Then we can break the q_{ij} integral into the sum of two parts, a circular and an elliptical one, and evaluate

$$q_{ij}^{(\text{circ})} = \int \text{d}\theta_x \text{d}\theta_y (f + 2\kappa_{\text{wl}}\theta^2 f_{,\theta^2})\theta_i\theta_j, \quad (14.66)$$

$$q_{ij}^{(\text{ell})} = -2 \int \text{d}\theta_x \text{d}\theta_y (\gamma_1\theta_x^2 - \gamma_1\theta_y^2 + 2\gamma_2\theta_x\theta_y)f_{,\theta^2}\theta_i\theta_j. \quad (14.67)$$

The result at first-order is (see problem 14.4)

$$\varepsilon_1 \approx 2\gamma_1, \quad \varepsilon_2 \approx 2\gamma_2. \quad (14.68)$$

Hence we can estimate $\gamma_{1,2}$ by measuring $\varepsilon_{1,2}$. The distortion components $\gamma_{1,2}$ are then calculated for each galaxy in the field, thereby creating an ellipticity map on the sky. Just as for the CMB map we can then estimate its power spectrum and compare it with the theoretical prediction. Naturally the measured ellipticity will be the sum of an intrinsic or “noise” galaxy ellipticity and the weak-lensing-induced ellipticity, $\gamma_i = \gamma_i^{(\text{N})} + \gamma_i^{(\text{WL})}$. Assuming the two components to be uncorrelated, the power spectrum is just the sum of the intrinsic and cosmic spectra. The intrinsic spectrum can be derived from the same shot noise spectrum we have derived earlier in Eq. (3.36) by substituting the weight w_i with the average intrinsic ellipticity $\langle w_i^2 \rangle = \langle (\gamma_1^{(\text{N})})^2 \rangle = \langle (\gamma_2^{(\text{N})})^2 \rangle \equiv \gamma_{\text{int}}^2$. If there are N sources in a volume V the intrinsic power is given by

$$P_{\text{int}} = \gamma_{\text{int}}^2 \frac{V}{N}. \quad (14.69)$$

A realistic value is $\gamma_{\text{int}} = 0.22$ [584].

So far we have calculated the distortion of the image of sources at a given comoving distance r . Since we are in the linear regime, we can add up all the transformation matrices for many sources at different distance. If $n(r)\text{d}r$ is the

number of sources (i.e. galaxies) in a shell dr with the normalization $\int_0^\infty n(r)dr = 1$, we can write the full transformation matrix D_{ij} [see Eq. (4.245)] as

$$D_{ij} = \int_0^\infty n(r') dr' \int_0^{r'} dr \left(1 - \frac{r}{r'}\right) r \psi_{,ij} = \int_0^\infty dr w(r) \psi_{,ij}, \quad (14.70)$$

where

$$w(r) \equiv \int_r^\infty dr' \left(1 - \frac{r}{r'}\right) r n(r'). \quad (14.71)$$

In the last step of Eq. (14.70) we have changed the order of integration by using the identity

$$\int_0^\infty dx f(x) \int_0^x dy g(x, y) = \int_0^\infty dy \int_y^\infty dx g(x, y) f(x). \quad (14.72)$$

All the integrations in r can be written as integrations in terms of z by the relation $dr = dz/H(z)$. In a non-flat space the coordinate r has to be replaced by χ as defined in Eq. (2.54). Therefore we obtain finally

$$D_{ij} = \int_0^\infty \frac{dz}{H(z)} w(z) \psi_{,ij}[\theta_x r(z), \theta_y r(z), r(z)]. \quad (14.73)$$

Now we need to use a theorem that projects along the z -axis the 3-dimensional power spectrum of a fluctuation field into a two-dimensional power spectrum: the Fourier-space Limber equation. If we have a field $f(x, y, r)$ projected along the r -direction with some unit-normalized weight $w(r)$, i.e.

$$F(\theta_x, \theta_y) = \int_0^\infty dr w(r) f(\theta_x r, \theta_y r, r), \quad (14.74)$$

then the two-dimensional power spectrum of F is given by [585]

$$P(q) = \int_0^\infty dr \frac{w(r)^2}{r^2} p\left(\frac{q}{r}\right), \quad (14.75)$$

provided that $p(k)$ is the power spectrum of f , and where q is the modulus of $\mathbf{q} = (q_1, q_2)$. See problem 14.5 for the derivation of Eq. (14.75). Now, the distortion field (14.70) is indeed such a projected field so that we can write

$$\kappa_{wl} = -\frac{1}{2}(D_{11} + D_{22}) = -\frac{1}{2} \int_0^\infty dr w(r) \psi_{,ii}, \quad (14.76)$$

(sum over i). Applying Limber's theorem we obtain the power spectrum of the convergence field κ_{wl} :

$$P_{\kappa_{wl}}(q) = \frac{1}{4} \int_0^\infty dr \frac{w(r)^2}{r^2} P_{\psi,ii}\left(\frac{q}{r}\right) = \frac{1}{4} \int_0^\infty dz \frac{W(z)^2}{H(z)} P_{\psi,ii}\left(\frac{q}{r}\right), \quad (14.77)$$

where $W(z) \equiv w[r(z)]/r(z)$. What is the spectrum of $\psi_{,ij}$? In Fourier space we have $\psi \rightarrow \psi_k e^{ik \cdot x}$ and $\psi_{,ij} \rightarrow -k_i k_j \psi_k$, so that the diagonal component is given by

$$P_{\psi,ii} = k^4 P_{\psi_k}. \quad (14.78)$$

If there is no anisotropic stress then $\psi = \Phi - \Psi = 2\Phi$ and from the Poisson equation for Φ one obtains

$$k^2 \psi_k = 3a^2 H^2 \Omega_m \delta_m, \quad (14.79)$$

where we have assumed that the fluctuations are only produced by the component with subscript m and that gravity is described by General Relativity. Then we can write

$$P_{\psi,ii} = k^4 P_{\psi_k} = 9H^4 \Omega_m^2 (1+z)^{-4} P_{\delta_m}. \quad (14.80)$$

Finally, by using Eqs. (14.77) and (14.80), we find that the power spectrum of the convergence κ_{wl} is [585, 586]

$$P_{\kappa_{wl}}(q) = \frac{9H_0^3}{4} \int_0^\infty dz \frac{W(z)^2 E^3(z) \Omega_m^2(z)}{(1+z)^4} P_{\delta_m} \left(\frac{q}{r(z)} \right), \quad (14.81)$$

where $E(z) = H(z)/H_0$ and

$$W(z) = \int_z^\infty \frac{d\tilde{z}}{H(\tilde{z})} \left[1 - \frac{r(z)}{r(\tilde{z})} \right] n[r(\tilde{z})]. \quad (14.82)$$

For small angles (large q) one can also write $q = \ell/\pi$ and estimate the spectrum as a function of the approximate multipole ℓ . If the m component is pressureless and uncoupled matter, then the expression can be simplified by noting that $E^4(z) \Omega_m^2/(1+z)^4 = (\Omega_m^{(0)})^2 (1+z)^2$. Notice that the selection function $n(z)$ is often directly given (normalized so that $\int n(z) dz = 1$) instead of $n(r)$. In this case one has to remember that $n(z) dz = n(r) dr$ and therefore $n[r(z)] = n(z) H(z)$. A typical observational distribution function is often parametrized as

$$n(z; z_0, \alpha) = z^2 \exp \left[-(z/z_0)^\alpha \right], \quad (14.83)$$

where α is a number of order unity that is fixed by observations. In the region $z \ll 1$ one has $n(z) dz \sim z^2 dz$, so that we are sampling all (or a constant fraction of) the galaxies in the spherical volume whose radius grows as $\propto z$. The decrease of $n(z)$ in the limit $z \rightarrow \infty$ takes into account that our survey is missing galaxies because they are too faint and red-shifted.

For the other components $\psi_{,ij}$ instead of Eq. (14.80) we have

$$P_{\psi,ij} = k_i k_j k_m k_n P_\psi = 9H^4 \Omega_m^2 (1+z)^{-4} \frac{k_i k_j k_m k_n}{k^4} P_{\delta_m}. \quad (14.84)$$

The general form of the power spectrum for D_{ij} is then

$$P_{ijmn}(q) = \frac{4k_i k_j k_m k_n}{k^4} P_{\kappa_{wl}}(q). \quad (14.85)$$

Now we have $\gamma_1 = (D_{22} - D_{11})/2$ and $\gamma_2 = -D_{12}$ thus

$$P_{\gamma_1} = (P_{2222} + P_{1111} - 2P_{1122})/4 = c_1 P_{\kappa_{wl}}, \quad (14.86)$$

$$P_{\gamma_2} = P_{1122} = c_2 P_{\kappa_{wl}}, \quad (14.87)$$

where $c_1 = (k_x^2 - k_y^2)^2/(4k^4)$ and $c_2 = k_x^2 k_y^2/k^4$.

Since the shear spectra are proportional to the convergence spectrum (at linear order) one can build combinations of the shear that are equal to $P_{\kappa_{wl}}$ and also combinations which are zero, as e.g., $(c_2 P_{\gamma_1} \pm c_1 P_{\gamma_2})/(2c_1 c_2)$. Defining the so-called *electric* and *magnetic* shear components as

$$E = \cos(2\phi)\gamma_1 + \sin(2\phi)\gamma_2, \quad (14.88)$$

$$B = -\sin(2\phi)\gamma_1 + \cos(2\phi)\gamma_2, \quad (14.89)$$

where the angle ϕ is the polar angle defined by the relation $\mathbf{k} = \{k \cos(\phi), k \sin(\phi)\}$, we find that

$$P_E = P_{\kappa_{wl}}, \quad (14.90)$$

$$P_B = 0. \quad (14.91)$$

Therefore, at first order the convergence power spectrum is all we need to test for cosmology with weak lensing. The power spectrum of the magnetic part B of the shear field, not excited by scalar perturbations, can be used as an observational test for consistency.

We can easily generalize the convergence power spectrum to the case in which we correlate sources in a redshift bin centered around z_i with sources in a bin around z_j [587]. In this case one sees that the spectrum (14.81) becomes

$$P_{ij}(\ell) = \frac{9H_0^3}{4} \int_0^\infty dz \frac{W_i(z)W_j(z)E^3(z)\Omega_m^2(z)}{(1+z)^4} P_{\delta_m} \left(\frac{\ell}{\pi r(z)} \right), \quad (14.92)$$

where

$$W_i(z) = \int_z^\infty \frac{d\tilde{z}}{H(\tilde{z})} \left[1 - \frac{r(z)}{r(\tilde{z})} \right] n_i[r(\tilde{z})]. \quad (14.93)$$

Now the distribution $n_i(z)$ will be non-zero only inside the i -th redshift bin and for small bins $W_i(z)$ may be approximated by $\Delta z [1 - r(z)/r(z_i)] n_i/H(z_i)$ for $z < z_i$ and zero outside. The weak lensing window function is always a broad function of z . The cosmological information is then less “localized” in z than in other methods (e.g., supernovae or clustering).

The expression above for the convergence power spectrum holds only when we can neglect the non-linear distortions of P_{δ_m} . However, this would be acceptable only up to relatively small ℓ 's, say up to $\ell < 500$ at most. Discarding all the information above would weaken a lot the constraints on the cosmological parameters. It is then necessary to employ the non-linear corrections to the power spectrum we have seen in Section 12.5. The spectrum P_{δ_m} will then be replaced by its corrected version that we can schematically indicate with $P_{NL}(P_\delta)$. Of course then the problem arises of calibrating the non-linear corrections with extensive N -body simulations.

The estimate of the *matter* power spectrum via cosmic shear can be compared directly with the direct estimate of the *galaxy* spectrum to determine the bias. This is more effectively achieved by cross-correlating the galaxy density field and the weak lensing convergence field [588]. Further information is contained in higher-order correlations of ellipticities induced by the non-Gaussianity of the matter distribution.

We have derived in Section 13.4 the Fisher matrix for the power spectrum. Since in Eq. (14.92) P_{ij} is a linear function of the power spectrum P_{δ_m} , one can derive a similar expression for the weak lensing case. In general, instead of calculating the spectrum at all ℓ 's (which is computationally demanding), we can calculate it at some interval $\Delta\ell$ and then linearly interpolate, considering that there are $(2\ell + 1)$ modes per multipole ℓ . If the survey covers a fraction f_{sky} of the full 4π sky, then only such fraction of modes are measurable independently. In this case the final result for a survey is (for the full derivation, see Refs. [589, 590]):

$$F_{\alpha\beta} = f_{\text{sky}} \sum_{\ell} \frac{(2\ell + 1)\Delta\ell}{2} \frac{\partial P_{ij}(\ell)}{\partial p_{\alpha}} C_{jk}^{-1} \frac{\partial P_{km}(\ell)}{\partial p_{\beta}} C_{mi}^{-1}, \quad (14.94)$$

which is summed over repeated indices (calculated as usual on the fiducial values). This is indeed of the form we have seen in Eq. (13.73). Here the cosmological parameters are p_{α} and the covariance matrix is given by

$$C_{jk} = P_{jk} + \delta_{jk} \gamma_{\text{int}}^2 n_j^{-1}, \quad (14.95)$$

where the second term is the intrinsic ellipticity term (14.69) and n_j is the number of galaxies per steradians belonging to the i -th bin. We can also write

$$n_j = 3600 \left(\frac{180}{\pi} \right)^2 N f_j, \quad (14.96)$$

where N is the total number of galaxies per square arcminute and f_j is the fraction belonging to the j -th bin.

If there is anisotropic stress and the Poisson equation is not standard, the lensing potential is given by Eq. (11.204) in Section 11.6, i.e.

$$k^2\psi = 3a^2 H^2 \Omega_m \delta_m \Sigma, \quad (14.97)$$

where Σ is defined in Eq. (11.203). Then in this case the convergence spectrum is given by

$$P_{ij}(\ell) = \frac{9H_0^3}{4} \int_0^\infty dz \frac{W_i(z)W_j(z)E^3(z)\Omega_m^2(z)}{(1+z)^4} \Sigma^2 P_{\delta_m} \left(\frac{\ell}{\pi r(z)} \right), \quad (14.98)$$

where $\Sigma = q(1 - \zeta/2)$ in general will depend on z and ℓ . However this neglects the non-linear correction $P_{NL}(P_{\delta_m})$ that we have mentioned above. A full non-linear correction for general q , ζ does not exist so far. In this case one could assume as a first approximation that applying the non-linear correction to $[q(1 - \zeta/2)]^2 P_{\delta_m}$ would work, but this should be further tested with N -body simulations. Notice that the matter linear power spectrum P_{δ_m} depends also on the linear growth function which itself depends on the functions $q(k, z)$ and $\zeta(k, z)$.

Beside the problem of the non-linear correction, there are several sources of systematic uncertainty that can affect the scientific outcome of cosmic shear surveys. We have assumed the intrinsic galaxy ellipticity to be completely uncorrelated with the cosmic signal, but this is clearly an oversimplification. Nearby galaxies tend to align due to tidal effects. In general galaxies respond to the average gravitational field by orienting themselves toward concentrations, thereby inducing an effect which is opposite to the cosmic shear in which galaxies are distorted tangentially with respect to the potential gradient. Moreover, if the redshift is estimated photometrically, the z -bins will be estimated with some error and will in fact overlap. All this lies of course on top of the technical challenge of estimating the ellipticity of tens of thousands, and soon hundreds of millions, of faint tiny objects, which requires extremely precise knowledge of the telescope point spread function. A good example of a work dealing with many of these important details and how to control their effect is given in Ref. [591].

So far, the existing cosmic shear estimations have provided interesting and complementary constraints on $\Omega_m^{(0)}$ and on the spectrum amplitude σ_8 (see Fig. 14.8) but not yet directly on dark energy parameters, due to the limited depth. Several future projects, from ground to space, promise to fully exploit the cosmological potential of weak lensing.

14.5 Cluster abundances and baryon fraction

We have seen in Section 12.4 that analytical or numerical methods can estimate the number of objects that form at any given redshift, i.e. the mass function $n(M, z)$.

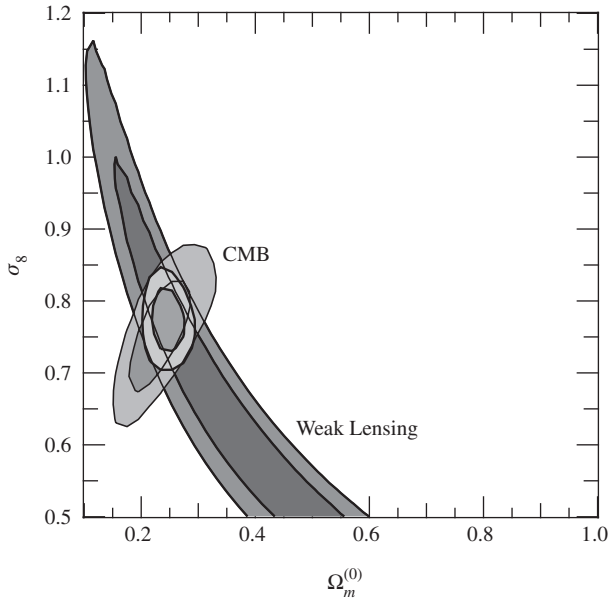


Figure 14.8 Constraints on $\Omega_m^{(0)}$ and σ_8 from the weak lensing observation (CFHTLS [592]) and the CMB observation (WMAP3 [14]) adopting a Λ CDM model. The combined constraints are indicated by the central region, demonstrating complementarity. From Ref. [583].

This quantity has a strong dependence on the *linear mass* variance $\sigma_M(z)$, obtained by filtering the fluctuation field with cells of size R that contains the mass M . It is important to stress the two words in *italic* in the previous sentence. The fact that we can use linear theory and that we refer to the total mass, not the luminous fraction, is the key to the interest in the mass function as a test of dark energy. The variance σ_M , in turn, depends directly on all the parameters that determine the mass power spectrum and its growth rate: $\Omega_m^{(0)}$, $\Omega_{DE}^{(0)}$, $w_{DE}(z)$, ... etc. Since the cell size is $R \propto [M/(\Omega_m^{(0)}\rho_c)]^{1/3}$, where ρ_c is the critical density, we expect the mass abundance at a particular scale M and redshift to be mostly sensitive to some combination of σ_8 (the power spectrum normalization) and $\Omega_m^{(0)}$ (which fixes the filter scale). It turns out from numerical estimates and from a comparison to N -body simulation that this relation can be conveniently written in the form (see e.g., Ref. [593])

$$\sigma_8(\Omega_m^{(0)})^\alpha \approx \beta, \quad (14.99)$$

with values $\alpha \approx \beta \approx 0.5 \pm 0.1$. Evaluating this relation at several masses and redshifts gives a precious independent test of cosmology.

The main difficulty in carrying out such a program is the determination of the mass of astrophysical objects. In principle we can use galaxies and clusters of any mass as target to estimate $n(M, z)$. However the most reliable application so far involves clusters of galaxies, mostly because their large scales bring them closer to linearity and because their masses can be estimated by more numerous and more robust methods than for galaxies.

There are indeed at least three independent methods to determine the mass of clusters: (i) hydrostatic equilibrium between the intra-cluster medium (ICM) and the gravitational potential, (ii) dynamics of member galaxies, and (iii) lensing. Among them, the first one has already provided some interesting results and it also looks very promising in the future.

Hydrostatic equilibrium for the ICM gas means that the gradient of the pressure P_{gas} equals the gravitational force:

$$\nabla P_{\text{gas}} = -\rho_{\text{gas}} \nabla \Phi_N, \quad (14.100)$$

where ρ_{gas} is the density and Φ_N is the gravitational potential. Assuming spherical symmetry we obtain

$$\frac{dP_{\text{gas}}}{dr} = -\frac{GM\rho_{\text{gas}}}{r^2}, \quad (14.101)$$

where we have used $\Phi_N = -GM/r$. Assuming the ideal gas equation of state [594]

$$P_{\text{gas}} = \frac{k_B \rho_{\text{gas}} T}{\mu m_p}, \quad (14.102)$$

where k_B is Boltzmann's constant and $\mu \approx 0.6$ is the mean molecular weight for a gas with the expected primordial composition and m_p is the proton mass, we obtain for the mass within a radius r :

$$M(r) = -\frac{r}{G} \frac{k_B T}{\mu m_p} \left(\frac{d \ln \rho_{\text{gas}}}{d \ln r} + \frac{d \ln T}{d \ln r} \right). \quad (14.103)$$

This provides a relation between the gas temperature T , the density profile ρ_{gas} , and the total cluster mass profile $M(r)$. In turn, the gas temperature can be estimated by comparing the X -ray bremsstrahlung emission with plasma models. The gas density profile is often parametrized by the so-called β -model distribution [595]

$$\rho_{\text{gas}} = \frac{\rho_0}{[1 + (r/r_c)^2]^{3\beta/2}}, \quad (14.104)$$

where $\beta = \mu m_p \sigma_r^2 / (k_B T)$ is the ratio of the gas kinetic energy (σ_r is the line-of-sight velocity dispersion) to temperature. If, in addition, the temperature gradient

$d \ln T / d \ln r$ is negligible (isothermal distribution) then the mass–temperature (M–T) relation reduces to

$$M(r) = \frac{3\beta k_B T(r)}{G\mu m_p} \frac{r^3}{r_c^2 + r^2} \approx (1.1 \times 10^{14} h^{-1} M_\odot) \beta \frac{T(r)r^3}{r_c^2 + r^2}, \quad (14.105)$$

where in the last expression r and T are in units of h^{-1} Mpc and keV, respectively. Although β is in principle measurable, it is always left as a free parameter in order to take into account at some level departures from the various assumptions (spherical model, ideal gas equation of state, isothermal distribution, etc.).

More complicated, and hopefully more realistic, models for the M–T relation have been proposed (see e.g., Ref. [596]). Using such mass–temperature relations the mass of several clusters has been established, for instance, by the satellites Chandra and XMM-Newton. Averaging over many clusters it is also possible to fit a universal simple mass–temperature relation. The simple fit provided by Ref. [596] is

$$M = M_5 \left(\frac{T}{5 \text{ keV}} \right)^\alpha, \quad (14.106)$$

with $\alpha \approx 1.5$ – 1.6 and $M_5 \approx 10^{14} M_\odot$. A value $\alpha = 3/2$ is indeed predicted for a virialized cluster, since in this case the velocity V_{vir} scales as $M^{1/3}$ and the gas kinetic energy is proportional to the temperature, so that $V_{\text{vir}}^2 \propto M^{2/3} \propto T$. Ultimately, a calibration of the mass–temperature relation will be provided by lensing mass estimations [597]. Once one has a well-calibrated M–T relation, it is possible to infer the cluster masses directly by measuring the temperature of the hot gas through a comparison of their X -ray spectra to plasma models. From the mass of several clusters one can finally reconstruct the mass function and compare it to the theoretical prediction we have seen in Section 12.4.

Using such methods, the following result for the relation $\sigma_8(\Omega_m^{(0)})^\alpha \approx \beta$ was found in Ref. [478]

$$\alpha = 0.21 - 0.22w_{\text{DE}} + 0.33\Omega_m^{(0)} + 0.25\Theta, \quad (14.107)$$

$$\beta = 0.5 - 0.1\Theta, \quad (14.108)$$

$$\Theta = (n_s - 1) + (h - 0.65), \quad (14.109)$$

for flat, $w_{\text{DE}} = \text{constant}$ models. Combining the cluster counts test with the other cosmological probes (e.g., CMB, SN Ia), it is possible to break the degeneracy between σ_8 , $\Omega_m^{(0)}$ and also to add other parameters to the analysis, e.g., the growth rate.

Clusters can contribute to constrain dark energy parameters in another way, first proposed by Sasaki [598] and Pen [599], expanding over previous work. As we

have seen for the supernovae, what is needed for cosmology is not necessarily a standard candle but rather a *standardizable* candle, i.e. a source whose absolute luminosity depends in a known way on an independent observable. If in clusters the mass of baryons that emit light, either X -ray emitting hot intracluster gas or optical galaxies, is a fixed universal fraction of the total mass, then by estimating the total mass we can estimate the total baryon mass and the total luminosity. This works just as for the supernovae: there, we estimate the total luminosity correlating it with the light-curve width; here, we correlate it with the total mass. In both cases we do not need to know the value of the absolute luminosity but only that it is constant or varies in a controlled way.

In clusters most of the baryons are actually in the intra-cluster medium, so for sake of simplicity we only consider the X luminosity. The fundamental assumption is that

$$\frac{M_{\text{gas}}}{M_{\text{tot}}} = \frac{\Omega_b}{\Omega_m} = \text{constant}, \quad (14.110)$$

for all clusters. This is indeed likely because clusters are very large: to make up their mass, one has to pile up all the matter in a radius of roughly 10 Mpc. It is difficult to imagine such large volumes containing wildly varying proportions of baryons and dark matter. There would simply be no time for any reasonable process to segregate matter on such large scales. Of course this ratio would not be constant for models that assume a species-dependent interaction like the ones we explored in Section 11.3.

So at least in standard cosmology, one expects all clusters to contain the fixed ratio of baryons to total matter set by cosmology. Now, the X -ray thermal bremsstrahlung luminosity L_X that comes from those baryons is proportional to the volume $V \propto r^3$ of the emitting region and to the square of the electron density ρ_e , i.e. to $\rho_e^2 r^3$. Since the mass M_{gas} is in proportion to $\rho_e V$, it follows that $L_X \propto M_{\text{gas}}^2 / r^3$ or $M_{\text{gas}} \propto (L_X r^3)^{1/2}$. We also notice that the X -luminosity is measured by an observed flux $\mathcal{F}_X = L_X / (4\pi d_L^2)$ [d_L is the luminosity distance defined in Eq. (2.63)], so we can also write $M_{\text{gas}} \propto d_L r^{3/2}$. On the other hand, from the hydrostatic equilibrium condition (14.103), we deduce that the total mass is $M_{\text{tot}}(r) \propto r$, if we assume an isothermal distribution and that $d \ln \rho_{\text{gas}} / d \ln r$ depends weakly on r (which is true for instance for all power-law $\rho_{\text{gas}} \sim r^n$). So finally we have

$$\frac{M_{\text{gas}}}{M_{\text{tot}}} \propto \frac{d_L r^{3/2}}{r} \propto d_L r^{1/2}. \quad (14.111)$$

There is a final step to make. The size r of the emission region is seen under the angle $\theta = r/d_A$ (d_A is the angular diameter distance) and therefore the gas fraction

within a fixed angle θ scales as

$$f_{\text{gas}} = \frac{M_{\text{gas}}}{M_{\text{tot}}} \bigg|_{<\theta} = A_1 d_L r^{1/2} = A_2 d_L d_A^{1/2} = A_3 d_A^{3/2}, \quad (14.112)$$

where A_1, A_2, A_3 are constants. Note that we have used the reciprocity relation (2.74) in the last step and absorbed the $(1+z)^2$ factor in A_3 . All the other constants of proportionality in the above relations are collected in the A_i 's factors. These factors contain a lot of interesting physics but no cosmological parameters, so we are not concerned with them here (see e.g., Refs. [598, 599]). Then we see that $f_{\text{gas}} d_A^{-3/2}$ is independent of cosmological parameters. If we have a reference cosmology (superscript “ref”) and another cosmology (2), we have

$$f_{\text{gas}}^{(\text{ref})} = f_{\text{gas}}^{(2)} (d_A^{(\text{ref})}/d_A^{(2)})^{3/2}. \quad (14.113)$$

If (2) is the true cosmology then $f_{\text{gas}}^{(2)} = \Omega_b/\Omega_m$ and we can finally say that the predicted gas fraction obtained by using a reference cosmology is

$$f_{\text{gas}}^{(\text{ref})} = \frac{\Omega_b}{\Omega_m} \left(\frac{d_A^{(\text{ref})}}{d_A} \right)^{3/2}. \quad (14.114)$$

Fitting $f_{\text{gas}}^{(\text{ref})}$ to the real data obtained by converting X-ray flux and temperature within the angle θ , we can constrain the cosmological parameters in d_A .

As you have presumably expected by now, it is time for the caveats. The simple f_{gas} prediction above relies on many things, from hydrostatic equilibrium to universal composition. Some approximations are easy to improve. For instance we can take into account the baryons contained in the galaxies rather than in the ICM. Other effects can be estimated from N -body, such as the typical departure from hydro-equilibrium or from universal composition. For instance, Ref. [600] found a depletion parameter (the fraction of baryons that are thermalized within the cluster potential) near 87%. Some other uncertainties can be marginalized over in the likelihood. Allowing for considerable freedom in parametrizing these effects, a table of constraints on various cosmological parameters has been derived in Ref. [601] from 42 clusters observed by the Chandra X-ray satellite. The constraint from the f_{gas} test alone gives $w_{\text{DE}} = -1.14 \pm 0.31$, for flat space and constant w_{DE} (all results here and below are at 1σ). The results in Ref. [602] on a different cluster dataset give $\Omega_m^{(0)} = 0.32_{-0.05}^{+0.04}$ and $w_{\text{DE}} = -1.1_{-0.45}^{+0.60}$. In combination with SN Ia and CMB, the constraint of Ref. [601] tightens to $w_{\text{DE}} = -0.98 \pm 0.07$ (see Fig. 14.9).

Ref. [603] combined the f_{gas} test with the cluster abundance to derive simultaneous constraints on the equation of state and the growth rate parameter γ (see Section 11.1): $w_{\text{DE}} = -0.927_{-0.074}^{+0.066}$ and $\gamma = 0.44_{-0.15}^{+0.17}$ for the flat $w_{\text{DE}} = \text{constant}$ model.

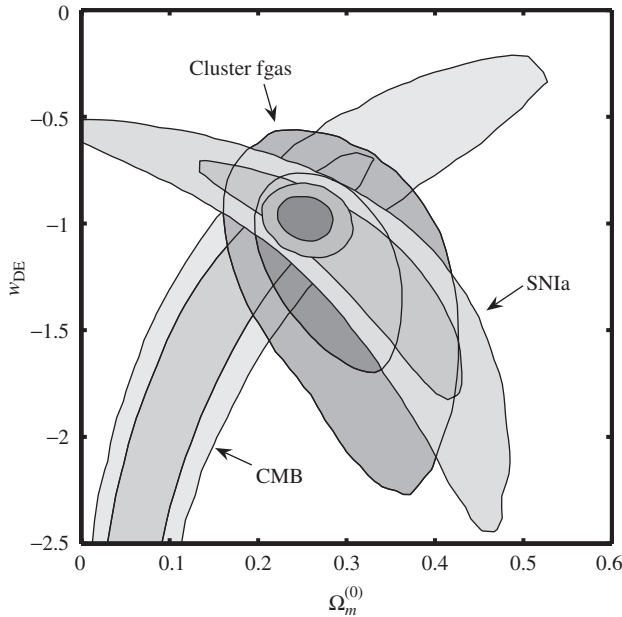


Figure 14.9 The 1σ and 2σ observational constraints on the parameters $\Omega_m^{(0)}$ and the constant equation of state w_{DE} of dark energy from the analysis of the Chandra f_{gas} data. We also show the bounds obtained from the CMB data [15] using a weak uniform prior on h ($0.2 < h < 2.0$) and SN Ia data [111]. The combined constraint from all three data sets is given by $\Omega_m^{(0)} = 0.253 \pm 0.021$ and $w_{\text{DE}} = -0.98 \pm 0.07$ at the 1σ confidence level. Adapted from Ref. [601].

14.6 Other probes

All astrophysical phenomena depend on the cosmological landscape on which they take place. A new and weighty ingredient of the cosmic recipe such as dark energy changes the landscape and, to various degrees, most of the phenomena we may observe. It is no wonder then that many ingenious tests of dark energy have been proposed. As we have emphasized several times, independent tests are precious even when they are not competitive, because they test possible systematic effects and help in identifying unknown biases.

The short list we propose below is certainly partial but gives a sample of the main promising ideas floating around.

14.6.1 Gamma-ray bursts

Gamma-ray bursts (GRBs) are powerful explosions produced by compact objects through several mechanisms. Long-duration bursts (from a few seconds to a few minutes) are probably generated by the collapse of a rapidly rotating, high-mass

star into a black hole. Since the radiation emission is so large ($\approx 10^{50}$ – 10^{54} erg), GRBs can be seen at very large distances, the current record being GRB 090423 at $z \approx 8.3$ [604], the most distant astrophysical object ever seen. The emission is probably highly collimated along the star rotation axis and the GRB flash is seen only if the axis is beamed towards the Earth.

Although the properties and physics of the GRBs are still very uncertain, several works have been devoted to finding correlations between their absolute luminosities and other observables so that they could be used as standard candles. Due to their typical high redshift, their use as candles would be extremely welcome to extend the reach of SN Ia.

There are several observables that could be used as luminosity correlators. These include the light-curve peak energy E_{pk} , the time-lag τ_{lag} between the arrival time of high- and low-energy photons, the variability of the light-curve (which in turn has been defined in several ways), the time t_{break} at which the afterglow has a break in power (which is attributed to the beaming of the emission), and the rise-time or duration of the light curve. Those observables can be correlated directly to the isotropic luminosity L_{iso} or to the beamed luminosity E_{γ} , i.e. the luminosity corrected for the beaming, which requires an estimate of the beam angular size through t_{break} . As we can expect, the use, validity, and interpretation of the various observables have generated much controversy that only further data can help to resolve (see Refs. [605, 606, 607] for the current status). So far, using only a few dozens of GRBs, the best correlations have been obtained using E_{pk} and E_{γ} , with significant improvement adding a second observable as t_{break} or the variability [608, 609]. The constraints on dark energy models are very preliminary and still very weak; they show general agreement with SN Ia.

14.6.2 Age tests

As we have seen in Section 5.1, the oldest stars we can observe in the globular clusters provided another independent test for the presence of dark energy. We recall that the Hubble parameter $H(z)$ is given by

$$H(z) = -\frac{1}{1+z} \frac{dz}{dt}. \quad (14.115)$$

This equation suggests that if we know the age difference Δt between two galaxies separated by the redshift Δz it is possible to obtain $H(z)$ directly [610, 611, 612]. The idea of using stellar ages to map the expansion rate is very appealing because it makes use of a set of assumptions totally different from the standard candles/rulers. Here it is the cosmic chronology that we are testing, not the geometry of light propagation. The difficulty of this method is to find accurate and readable clocks.

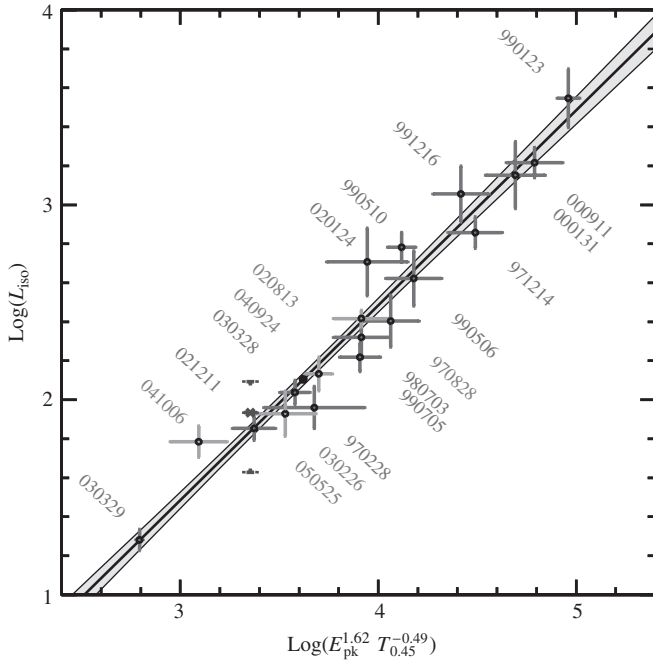


Figure 14.10 Correlation between GRB peak luminosity E_{pk} , beamed luminosity L_{iso} and a duration parameter (here called $T_{0.45}$) for 19 GRBs. The best fit and its 1σ uncertainty are represented by the solid line and by the shaded region, respectively. From Ref. [609].

Computing age *differences* avoids the problem of determining the absolute galaxy age that would require knowledge of the galaxy evolution *before* star formation, which cannot be estimated by methods based on stellar population evolution. Alternatively the pre-formation time delay can be marginalized over, although this implies that there is a single time delay for all systems.

Globular clusters are particularly useful since their low gravity halts the process of star generations and in many cases we can safely assume that most of their stars were born in a single star-burst episode. However, individual stars or globular clusters can be seen only in our Galaxy and in some nearby systems. Hence we have to use integrated signals to go further, i.e. the cumulative light of stellar populations. This of course introduces the problem of mixing different populations, i.e. young stars with old ones.

The method of galaxy age dating implies therefore a careful comparison of models of population synthesis, i.e. the numerical evolution of the integrated spectrum of stellar populations with the galaxy spectra. The present numerical codes are very sophisticated, including parameters that model metallicities and multiple star-burst

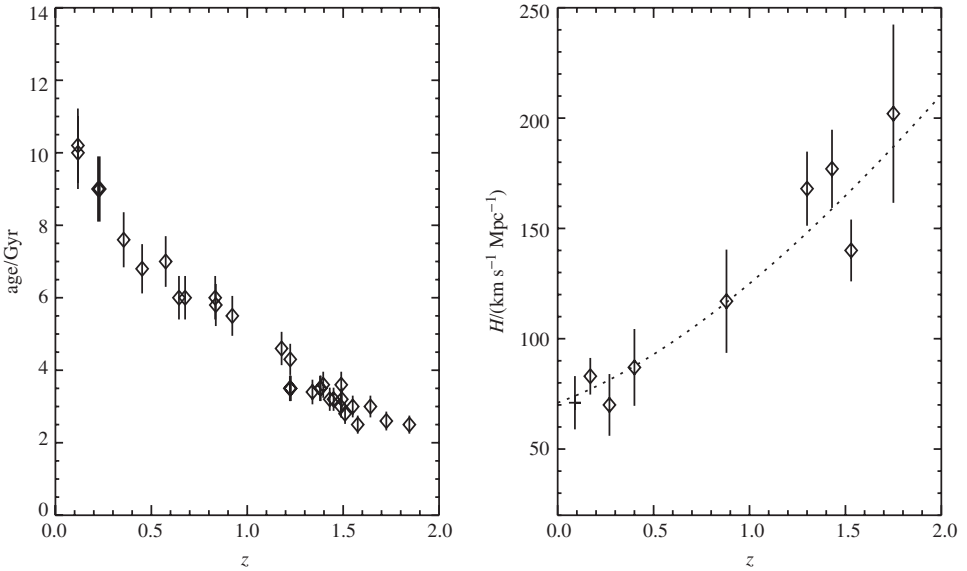


Figure 14.11 Left panel: absolute age for 32 passively evolving galaxies determined from fitting stellar population models as a function of the redshift z . Right panel: the value of the Hubble parameter versus z as derived from the differential ages of galaxies in the left panel. The dotted line is the value of $H(z)$ for the Λ CDM model. From Ref. [612].

episodes. It is reassuring that in most cases the results of independent codes agree to quite a surprising precision.

Combining galaxy ages from a sample of 32 early-type galaxies for the redshift $z < 1.8$ which are supposed to have been evolving passively (i.e. with very low star formation) with CMB observations, Ref. [612] finds that the Λ CDM model is consistent with the data. Using the same galaxy sample data, Ref. [613] finds the constraint $-1.21 < w_{\text{DE}} < -0.88$ (2σ) for the constant equation of state of dark energy.

14.6.3 Strong lensing

Just as we can estimate the abundance of collapsed objects via semi-analytical means, so we can estimate how much strong gravitational lensing we can expect from these objects. The advantage with respect to cluster and galaxy number counts is that lensing depends only on the lens gravitational potential and therefore we can use directly the mass power spectrum predicted by the models. The disadvantage is that the number of lensed sources depends on the detailed profile of the lens and there is therefore an extra layer of parametrization.

Strong lensing effects are of various kinds: one can consider the number of image multiples, of giant arcs, or even the statistics of time delay between variable

signals in multiple images. Here we limit ourselves to a short discussion of the expected probability of multiple images.

Let us then calculate how many multiple images we can expect in a survey as a function of cosmological parameters and of the survey properties. We want to estimate the probability $P(> \Delta\theta_0, z_s)$ that a source at redshift z_s is split into two (or more) images separated by an angle larger than $\Delta\theta_0$ by (the dark matter halos of) galaxies and clusters along the line of sight. If a halo of mass M at redshift z is sufficiently massive to split the source image by at least $\Delta\theta_0$ we associate with it a non-zero cross section $\sigma(M, z, \Delta\theta_0)$ equal to the area within which the lensing takes place. If in a space volume $dV = A dr$ of area A and comoving depth dr we have N lenses of mass M each offering a target area σ for a lensing event, and defining $n(M, z)dM$ as the number density of lenses of mass within $M, M + dM$ (i.e. the mass function), we have that the probability of a lensing event is $N\sigma/A$, i.e. [615]

$$dP(M, z) = \frac{N\sigma}{A} = \frac{(ndV)\sigma}{A} dM = n\sigma(dr)(dM) = n\sigma \frac{dr}{dz}(dz)(dM), \quad (14.116)$$

as the differential probability of a lensing event induced by lenses of mass M at redshift z . The mass function is normalized so that in a survey of volume V we have $V \int n(M)dM = 1$. Integrating over the masses we obtain

$$dP = \frac{dr}{dz} dz \int_0^\infty n(M, z)\sigma(M, z, \Delta\theta_0) dM. \quad (14.117)$$

The integrated probability of a lensing splitting larger than $\Delta\theta_0$ along the line of sight is then

$$P(> \Delta\theta_0) = \int_0^{z_s} dz \frac{dP(z, \Delta\theta_0)}{dz}. \quad (14.118)$$

This quantity can be directly compared to the observed number of lenses: if we take a sky survey of N sources at redshift z_s , we should see PN objects split by more than $\Delta\theta_0$. Let us estimate then $P(> \Delta\theta_0)$.

A background source S is split by an angle $\Delta\theta_0$ by a lens L if S lies within a radius $\xi_0(\Delta\theta_0)$ of the lens, called critical radius. This radius depends of course on the model of the lens itself, i.e. on its radial profile and on its sub-structure. The lens modeling is the main uncertainty in this kind of study since the only information on the dark halo profiles depends on N -body simulations and on lensing maps (which however can be performed only for some clusters). Let us assume here as a model for the lenses the simplest case, the singular isothermal sphere (SIS), characterized by a mass density $\rho \sim r^{-2}$. This profile might be a first rough approximation to the spherical halos of galaxies and clusters. The SIS produces in fact flat rotation curves and therefore could approximate the true halo at least in the outer regions.

It is customary to choose the normalization constant as

$$\rho(r) = \frac{\sigma_v^2}{2\pi Gr^2}, \quad (14.119)$$

where σ_v is the velocity dispersion of the stars (for the galaxies) or the member galaxies (for clusters). Typical values are of the order of 200–300 km/sec. We can relate the velocity dispersion to the halo mass by assuming that the virial theorem holds within a distance r_{200} that contains an average mass density 200 times the cosmic density (see the spherical collapse model in Section 12.3). Considering the cosmological evolution in z we can write therefore $\rho(r_{200}) = 200E^2(z)\rho_c^{(0)}$, where $E(z) = H(z)/H_0$. Then we have

$$\sigma_v^2 \approx \frac{GM}{2r_{200}}, \quad (14.120)$$

where M is the mass within r_{200} . Combining with Eq. (14.119) applied to r_{200} , we obtain the following explicit relation

$$M(z) \approx 0.656 \cdot 10^{15} h^{-1} M_\odot \left(\frac{\sigma_v}{1000 \text{ km/sec}} \right)^3 E^{-1}, \quad (14.121)$$

where $E(z)$ depends on the cosmological model ($1M_\odot \approx 2 \cdot 10^{33}$ grams).

The deflection of a beam passing through a SIS is discussed at length in textbooks [614], so we just quote the result [615]

$$\beta = 4\pi \left(\frac{\sigma_v}{c} \right)^2 \frac{d_{ls}}{d_s}, \quad (14.122)$$

where $d_{s,l,ls}$ are the angular diameter distances to the source, the lens, and between lens and source respectively.

Let us define the impact parameter b as the transverse distance on the lens plane between the unperturbed line of sight to S and the line of sight to l , i.e. $b = d_l \theta$, where θ is the angular separations between the two lines of sight. A splitting occurs only if $\theta < \beta$ that is if $b < \xi_0 \equiv d_l \beta$ and the splitting angle in arcminutes and in radians is simply

$$\Delta\theta = 2\beta \approx 0.96' \left(\frac{\sigma_v}{1000 \text{ km/sec}} \right)^2 \frac{d_{ls}}{d_s} \approx 3.7 \cdot 10^{-4} \left(\frac{M(z)E(z)}{M_{15}} \right)^{2/3} \frac{d_{ls}}{d_s}, \quad (14.123)$$

where $M_{15} \equiv 10^{15} h^{-1} M_\odot$. We can invert the last equivalence to obtain explicitly the mass M that is needed to produce a $\Delta\theta$ splitting:

$$M \equiv 0.7 \cdot 10^{15} h^{-1} M_\odot \left(\frac{\Delta\theta}{1'} \right)^{3/2} \left(\frac{d_s}{d_{ls}} \right)^{3/2} E(z)^{-1}. \quad (14.124)$$

Then the cross section for splitting $\Delta\theta$ equal to or larger than $\Delta\theta_0$ is

$$\sigma = \pi \xi_0^2 = \frac{\pi}{4} d_l^2 (\Delta\theta)^2, \quad (14.125)$$

while it is zero (i.e. no lensing event) for $\Delta\theta < \Delta\theta_0$.

We can now put everything together into Eq. (14.118) and obtain (see Refs. [616, 617])

$$\begin{aligned} P(> \Delta\theta_0) &= \frac{\pi}{4} \int_0^{z_s} dz \frac{dr}{dz} d_l^2 \int_{M_0}^{\infty} n(M, z) (\Delta\theta)^2 dM \\ &= 1.36 \cdot 10^{-7} \frac{\pi}{4} \int_0^{z_s} dz \frac{E^{1/3} d_l^2 d_{ls}^2}{H_0 d_s^2} \int_{M_0}^{\infty} n(M, z) \left(\frac{M(z)}{M_{15}} \right)^{4/3} dM, \end{aligned} \quad (14.126)$$

where we have used $dr/dz = 1/(H_0 E(z))$, and M_0 is the mass that produces the split $\Delta\theta_0$. To evaluate the mass function $n(M, z)$ we can use the Press–Schechter approximation we discussed in Section 12.4. Notice that Eq. (14.126) is dimensionally correct since $n(M, z)dM$ has the dimension of the inverse of a volume.

A similar estimate of $P(> \Delta\theta_0)$ can be produced with more realistic lens models, for instance a generalized Navarro–Frenk–White profile [618, 619]

$$\rho_{\text{NFW}}(r) = \frac{\rho_s r_s^3}{r^\alpha (r + r_s)^{3-\alpha}}, \quad (14.127)$$

with $0 < \alpha < 3$. Here r_s is the core radius and ρ_s is a normalization constant. This profile fits halos over a large range of masses and scales and therefore can be used to improve upon the isothermal sphere.

Before we can compare P to the frequency of lenses in real surveys we need to consider that the lenses are also magnified and therefore we are likely to see more lenses than one would predict in a given survey to some limiting magnitude. This biasing depends on the lens model and on the source distribution in apparent magnitude and can lead to a number of observed lenses several times larger than predicted without the magnification effect [617].

The bias-corrected P_{obs} can then be compared to observations, i.e. to the number of lenses in catalogs like the Cosmic Lens All-Sky Survey (CLASS) [620] and SLACS [621, 622], whose total number does not exceed ~ 100 –150 confirmed lenses. Similar estimates of strong lensing probabilities can be applied to the statistics of giant arcs in the background of clusters comparing observations to N -body simulations [623, 624, 519, 625] and to time delays between multiple images [626].

The use of strong lens statistics to constrain dark energy will depend more and more on the availability of high resolution N -body simulations capable of taking into account cluster sub-structure. The analytical modeling we have discussed

above depends in fact strongly on the lens model and on various other parameters, from the lens redshift distribution to the magnification bias. For instance, changing the parameter α in Eq. (14.127) from 1 to 2 decreases the predicted number of lenses by four orders of magnitude. The present constraints on dark energy are consequently still very weak (see e.g., Ref. [627]).

14.6.4 The redshift drift

All the probes we have reviewed in this chapter deal with position, brightness, spectrum, or shape of astrophysical sources. This section is exceptional in that it makes use of the *time variation* of the above observables. What we mean is that we can observe in real time the variation induced by cosmic expansion on the properties of the sources. This general concept has been dubbed *real-time cosmology* [464].

The idea of measuring the cosmic expansion as it occurs sounds daunting and this was the conclusion in 1962 on the first paper ever to consider this possibility [628]. Sandage considered the change in redshift of distant galaxies after a suitable time span, say, 10 years. Since during that time the expansion rate $H(z)$ changes by some ΔH , the recession velocity of the source and its redshift should change accordingly. If the Universe decelerates (accelerates), the redshift z should decrease (increase). If the Universe goes through phases of deceleration and acceleration, as implied in the Λ CDM model, then one should be able to detect increments or decrements according to the distance. Sandage thought that the redshift drift would have produced a clear discrimination between steady-state and big bang cosmologies.

It is easy to derive an order of magnitude value. In a time span of Δt years, the dimensionless change $\Delta z/(1+z)$ will be near the dimensionless combination $H_0 \Delta t$. For $\Delta t = 10$ yrs we can expect an effect of the order of 10^{-9} . For a recession velocity near c this amounts to a change in velocity of a few centimeters per second per year. Daunting, indeed.

Many years after Sandage's proposal (and after other efforts at revitalizing the issue, e.g., [629, 630, 631]¹), Loeb [632] reconsidered the idea and realized that the current technology applied to the Lyman- α absorption lines in the quasar line of sights allows Doppler redshift measurements of a few meters per second, not that far from Sandage's effect. According to recent studies, the ultra-stable, high-resolution spectrograph of the European Extremely Large Telescope (EELT), an optical telescope with diameter in the range 40–50 m planned for the second or third decade of this century may achieve the required sensitivity [633]. The measure will

¹ It is interesting to note that Rüdiger in 1982 considered the possibility of using the redshift drift to test for the acceleration predicted by a Brans–Dicke model [631].

be realized by monitoring the shift of Lyman- α absorption lines in some dozens of distant quasars for a period of one or two decades. Lyman- α lines are known to be very stable (because they do not contain pointlike sources) and with low peculiar velocities (because the Lyman- α clouds reside mostly in underdense regions).

Let us derive the redshift variation in the FLRW metric. The observed redshift of a given source, which emitted its light at a time t_s , is today (i.e. at time t_0)

$$z_s(t_0) = \frac{a(t_0)}{a(t_s)} - 1. \quad (14.128)$$

After a time interval Δt_0 (Δt_s for the source) it becomes

$$z_s(t_0 + \Delta t_0) = \frac{a(t_0 + \Delta t_0)}{a(t_s + \Delta t_s)} - 1. \quad (14.129)$$

The observed redshift variation of the source is, then,

$$\Delta z_s = z_s(t_0 + \Delta t_0) - z_s(t_0) = \frac{a(t_0 + \Delta t_0)}{a(t_s + \Delta t_s)} - \frac{a(t_0)}{a(t_s)}. \quad (14.130)$$

This can be re-expressed, after an expansion at first order in $\Delta t/t$, as:

$$\Delta z_s \simeq \Delta t_0 \left(\frac{\dot{a}(t_0) - \dot{a}(t_s)}{a(t_s)} \right), \quad (14.131)$$

where we have used the relation $a(t_0)/a(t_s) = \Delta t_0/\Delta t_s$. In terms of the Hubble parameter $H = \dot{a}/a$, we finally obtain

$$\frac{\Delta z_s}{1 + z_s} = H_0 \Delta t_0 \left[1 - \frac{H(z_s)}{(1 + z_s)H_0} \right]. \quad (14.132)$$

The Sandage effect can then in principle map the expansion rate $H(z)$ directly, as we show in Fig. 14.12 [634, 635, 636]. In terms of an apparent velocity shift of the source, the result can be written as $\Delta v = c\Delta z_s/(1 + z_s)$. From Eq. (14.131) we find that the effect vanishes at all redshifts for a constant-velocity expansion $\dot{a} = \text{constant}$, for instance, a model with total equation of state $w_{\text{eff}} = -1/3$.

The Extremely Large Telescopes collaboration [633] has provided some detailed calculations of the expected error on the velocity shift Δv . They predict a typical error:

$$\sigma_{\Delta v} = 2 \left(\frac{2370}{S/N} \right) \left(\frac{30}{N_{\text{QSO}}} \right)^{1/2} \left(\frac{5}{1 + z_{\text{QSO}}} \right)^{1.7} \text{ cm/s}, \quad (14.133)$$

where S/N is the signal to noise ratio for pixels of $0.0125 \text{ \AA}$, N_{QSO} is the number of QSO's spectra observed, and z_{QSO} is their redshift.

Detecting the Sandage effect is extremely challenging, to put it mildly. Despite this, it has so many interesting features that it deserves to be pursued to the maximal extent. First, it would provide the first direct probe of the cosmological expansion.

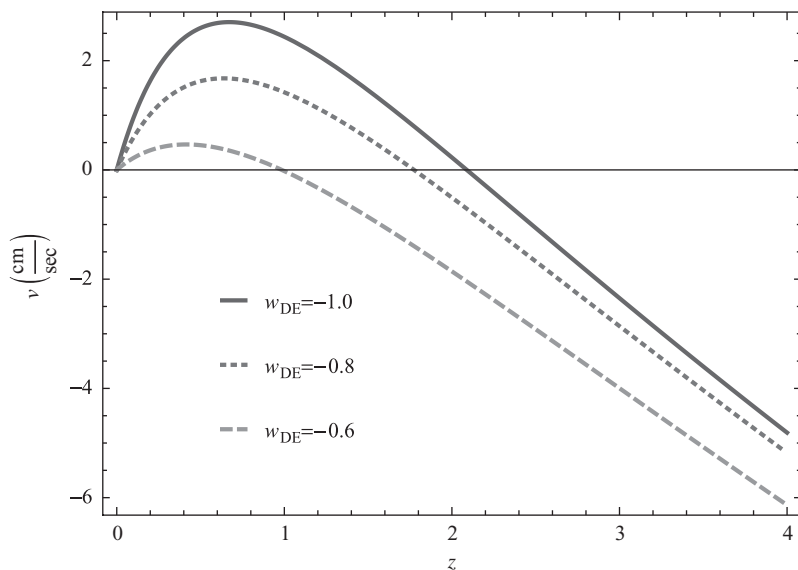


Figure 14.12 Sandage effect. Plot of the velocity shift after 10 years for three different values of the dark energy equation of state w_{DE} .

One must not forget that the whole of cosmology relies on the (extremely well grounded) assumption that the redshift is almost entirely due to the expansion. Second, it would be a completely independent test of acceleration. Third, its validity does not depend on calibration and extrapolation (unlike the supernovae test), nor on a complex modeling with many unknowns (such as e.g., the baryon acoustic oscillations or the cosmic shear). Fourth, the signal increases directly with the time span, while the statistical noise scales as $t^{-1/2}$, so that the signal-to-noise ratio increases as $t^{3/2}$. Time is a costly but virtually inexhaustible resource. Fifth, in principle it can be searched for not only in Lyman- α lines but in all extragalactic sources. Sixth, it can distinguish between a true acceleration and an apparent one induced by line-of-sight inhomogeneities as we have seen in Section 10.1.

The Sandage effect is just one example of real-time cosmology. Another possibility is to detect changes in the angular separation of distant sources [464]. Any anisotropic Universe (including some proposed to explain the cosmic acceleration, see Refs. [637, 455, 638] and Section 10.1) undergoes an expansion rate that depends on the source distance or direction. This in turn implies that the angular separation between any two sources will vary in time. Here again, assuming a time span of 10 years one can expect an effect of 10^{-9} . The good news is that this requires a purely astrometric measurement: taking as order of magnitude a value of 10^{-9} rad we obtain about 200 microarcseconds (μ as), which is easily within

current or near future capability. The satellite mission GAIA for instance has a goal of a few microseconds of arc for millions of stellar and extragalactic sources [639], so we could expect to be able to put strong constraints on anisotropy. The bad news is that the anisotropic models proposed to explain the cosmic acceleration are already pretty much constrained and allow only for a marginal anisotropy.

The parallax $\Delta\gamma$ induced by the Earth's own motion v_{pec} on distant sources in say 10 years is also of the order of a microarcsecond, or more exactly

$$\Delta\gamma = \left(\frac{v_{\text{pec}}}{500 \text{ km/s}} \right) \left(\frac{d_A}{1 \text{ Gpc}} \right)^{-1} \left(\frac{\Delta t}{10 \text{ years}} \right) \mu \text{ as}, \quad (14.134)$$

where d_A is the diameter distance of the source. This dipolar effect competes with the cosmic anisotropic signal. However it is interesting on its own because this could allow a determination of quasar distances completely independent of standard candles, rulers, or clocks [640] and with much less unknown systematics. Cosmic parallax effects could be within reach of future astrometric satellite experiments and, along with the radial component measured by the Sandage effect, could complete the three-dimensional picture of the cosmic flow.

14.6.5 Comparing luminosity and angular diameter distance

Finally, we conclude our review of dark energy probes by a test not of dark energy itself but of a fundamental assumption that underlies many of the tests seen above, namely that photons are not created nor destroyed during their ride across the Universe. Of course we know that photons are indeed propagating through a perturbed Universe, so they undergo deviations, scattering, gravitational redshifts, and absorptions. However, most of these effects leave some footprint that we can detect and use to correct for the perturbation. For example, we can use image lensing to detect photon deviations, we can see sources at different wavebands to detect phenomena of absorptions and reemission, we can compare photons from different directions to detect red or blue shifts, and so on. But how can we tell whether a photon emitted by some supernovae just disappeared along its run, maybe because due to some exotic physics it decayed into some particles we do not see or because it slipped into an extra dimension?

In fact, we can test for such effects by recalling that we measure light ray paths in two different ways, by measuring luminosity distance (i.e. using standard candles) and by measuring angular diameter distances (standard rulers). While the first changes if the flux we receive is depleted by some exotic photon physics or by unaccounted astrophysical attenuation, the second remains unaltered. If this is the case, then the reciprocity or Etherington relation (2.74) does not hold anymore. This

relation is true in any general metric theory of gravity, regardless of the background, if photon flux is conserved. Its violation is a test of cosmic transparency.

We can then test for the reciprocity relation by writing it as

$$d_L = (1 + z)^{2+\alpha} d_A. \quad (14.135)$$

Comparing the d_L distance measured by SN Ia with the d_A distance measured by, e.g., BAO, it is possible to place constraints on α . In fact, a discrepancy between the two measures was reported in Ref. [641]. The most updated result using supernovae and $H(z)$ from age tests gives however a result perfectly compatible with the reciprocity relation [642],

$$\alpha = -0.01^{+0.08}_{-0.09}. \quad (14.136)$$

So although there is no indication of a departure from standard physics, there is still considerable room for some surprise here.

14.7 Problems

14.1 Evaluate the ISW spectrum (14.29) analytically for standard gravity ($\Sigma = 1$) with $\tau_{\text{op}} = 0$ under the assumption of a power-law spectrum $P_\delta(k) = Ak^n$, a constant total equation of state $w_{\text{eff}} < -1/3$, and a total perturbation growth $D_t \sim a^p$ ($p > 0$).

14.2 Derive Eq. (14.52).

Hint: use Eq. (14.51) and expand

$$e^{ikr\mu_{kr}} = \sum_0^\infty (2\ell + 1) i^\ell \mathcal{P}_\ell(\mu_{kr}) j_\ell(kr), \quad (14.137)$$

where μ_{kr} denotes the cosine of the angle between $\mathbf{k}$ and $\mathbf{r}$. Make use also of Eqs. (17.8), (17.9), (17.10), and (17.14) in the mathematical Appendix. From Ref. [643].

14.3 Verify Eq. (14.60).

Hint: first integrate $\Xi(\sigma)$ by parts.

14.4 Derive Eq. (14.68).

14.5 Derive Eq. (14.75).

Conclusion and outlook

The expression “dark energy” was put for the first time in print in a paper by Huterer and Turner in 1998 [216], just a few months after the discovery of cosmic acceleration with supernovae Ia. In December 1998 the name was already featured in the *New York Times*. The idea of a smooth, invisible, accelerating component of the Universe is however much older than this. Beside the original suggestion of a cosmological constant by Einstein, already in the 1980s some scientists suggested that one way of solving the conflict between the inflationary prediction $\Omega^{(0)} = 1$ and the observations of clustered matter $\Omega_m^{(0)} \approx 0.3 \pm 0.1$ was to introduce a cosmological constant [644, 645, 646] or a slowly varying scalar field [30]. The idea was pursued at a low intensity for several years (e.g., [647]), until the supernovae boom. Since then, the number of papers with “dark energy” in the title has grown almost exponentially (see Fig. 15.1), achieving stabilization only around 2008, probably also because the variety of models on the market has rendered the expression “dark energy” no longer fully informative.

This bewildering variety points directly to the essence of the dark energy concept: after many years of research, dark energy is still a question, not an answer. It is the name we give to the fact that something, a very weighty something, is missing from our knowledge of the cosmic dynamics. The main teaching is that what we thought was a rather simple Universe, with essentially matter and radiation or if you wish matter-like and radiation-like particles, revealed itself not only more complicated (the cosmological constant is actually the least complicated form of energy one can think of) but certainly more unknown. We faced the fact that we have “direct” information only on a few snapshots of the cosmic evolution: the inflationary perturbations, the primordial nucleosynthesis, the Cosmic Microwave Background, the present galaxy distribution, and a few and sparse observations in between. Too few to tell the whole story.

The sober realization that we still have a lot to learn has opened up many avenues that perhaps would not have been thought of if everything did conform to our naive

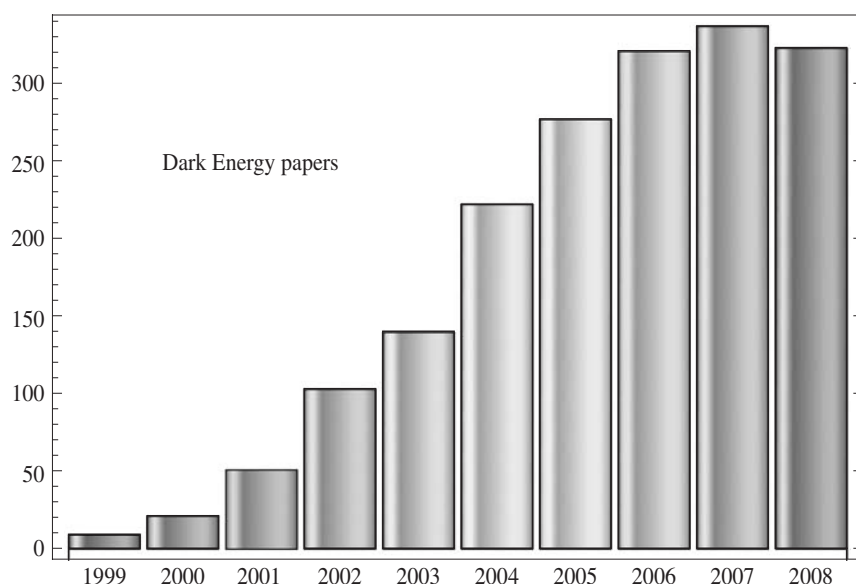


Figure 15.1 Papers with “dark energy” in the title from 1999 to 2008 in the SLAC Spires (www.slac.stanford.edu/spires/hep/).

expectations. The idea of seriously considering deviation from the Einstein gravity at large scales, rather than in the laboratory or in the solar system; the search for better probes of cosmic expansion and clustering at high redshift; the investigation of large-scale effects of inhomogeneities; many new ideas on how to exploit all astrophysical observables to complete our picture; all this research would probably have aroused much less interest if the dark energy mystery was not demanding it.

So this book is not as much about results as it is about suggestions and methods. Many models we have considered in turn might all be deadly wrong but they all teach us new ways of looking at the cosmological problem. The observations that institutions around the globe are proposing, financing, and carrying out are showing us the way to look further, harder, and deeper.

Where will the dark energy research go in the next years? The mere list of observational projects related to dark energy that is planned or underway is impressive [648]. The knowledge that will be acquired by these data will no doubt keep on constraining the classical cosmological parameters and maybe at some point will require something beyond a pure cosmological constant. This alone would probably classify as the most important discovery in cosmology since the Cosmic Microwave Background and the acceleration itself. If instead the error bars will keep focusing on the Λ spot, then the puzzles of coincidence and fine tuning will loom even more urgent over us. But on the way we can safely expect to reach several other goals: to probe gravity beyond the solar system, to reconstruct the clustering

evolution down to high redshifts, to map the total matter content over vast patches of sky, to confirm or confine the use of standard candles, rulers or clocks, to define to high precision the landscape on which astrophysical phenomena take place, to cross-check the many assumptions we currently make on our interpretations of the real data. Even if the nature of dark energy will continue to elude us, all the effort in this direction will not be in vain. It could as well happen that instead of a shorter route to the East we will find a whole new world.

16

Answers to the problems

Chapter 2

- 2.1** In the FLRW spacetime the energy-momentum tensor T_ν^μ is a function of the cosmic time t only. Then the conservation equation (2.90) yields

$$\nabla_\mu T_0^\mu = \dot{T}_0^0 + \Gamma_{0i}^i T_0^0 - \Gamma_{i0}^i T_i^i = 0, \quad (16.1)$$

where the Latin indices $i = 1, 2, 3$ are summed over. Since $T_0^0 = -\rho$, $T_j^i = P\delta_j^i$, and $\Gamma_{0j}^i = H\delta_j^i$, it follows that $\dot{\rho} + 3H(\rho + P) = 0$.

- 2.2** From Eqs. (2.42) and (2.43) the entropy densities of relativistic particles, for each spin state, are given by $s = 2\pi^2 T^3/45$ and $s = (7/8)2\pi^2 T^3/45$ for bosons and fermions, respectively. Before the annihilation of electrons and positrons, there were photons (2 spin states), 3 flavor neutrinos and anti-neutrinos (1 spin state for each), electrons (2 spin state) and positrons (2 spin state). Then the entropy before the annihilation with temperature T_1 and scale factor a_1 is

$$S(a_1) = s(a_1)a_1^3 = \frac{2\pi^2}{45} T_1^3 [2 + (7/8)(3 + 3 + 2 + 2)] a_1^3 = \frac{43}{90} \pi^2 T_1^3 a_1^3. \quad (16.2)$$

After the annihilation of electrons and positrons, neutrinos decoupled from the cosmic plasma so that the neutrino temperature T_ν and the photon temperature T_γ are different. Then the entropy after the annihilation is

$$S(a_2) = \frac{2\pi^2}{45} \left(2T_\gamma^3 + \frac{7}{8} \cdot 6T_\nu^3 \right) a_2^3. \quad (16.3)$$

The equality $S(a_1) = S(a_2)$ holds because of the entropy conservation. Using this equality together with the relation $a_1 T_1 = a_2 T_\nu$ for neutrinos, we obtain $T_\nu/T_\gamma = (4/11)^{1/3}$.

- 2.3** In the non-relativistic limit, $m^2 \gg p^2$, we have the approximate relation $E \simeq m + p^2/(2m)$ from the relation $E^2 = p^2 + m^2$. Plugging this into Eq. (2.40) it follows that

$$\rho \simeq \frac{g_*}{2\pi^2} m \exp[-(m - \mu)/T] \int_0^\infty p^2 \exp(-p^2/2mT) dp. \quad (16.4)$$

Using the integral $\int_0^\infty x^2 e^{-x^2} dx = \sqrt{\pi}/4$, we arrive at the density (2.48). Under the same non-relativistic approximation, the pressure (2.41) yields

$$P = \frac{g_*}{6\pi^2} \frac{1}{m} \exp[-(m - \mu)/T] \int_0^\infty p^4 \exp(-p^2/2mT) dp. \quad (16.5)$$

Using the integral $\int_0^\infty x^4 e^{-x^2} dx = 3\sqrt{\pi}/8$, the pressure (2.49) follows.

- 2.4** For the solution (2.89) the Hubble parameter is given by $H = n/(t_{\text{rip}} - t)$ with $n \equiv -2/[3(1 + w_{\text{DE}})] > 0$. Then the scalar curvature evolves as

$$R = 6(2H^2 + \dot{H}) = \frac{6n(2n + 1)}{(t_{\text{rip}} - t)^2}, \quad (16.6)$$

which diverges as $t \rightarrow t_{\text{rip}}$. Under the approximation $H_0 \simeq n/(t_{\text{rip}} - t_0)$, which corresponds to neglecting the contribution of non-relativistic matter today, it follows that

$$t_{\text{rip}} - t_0 \simeq \frac{2}{3|1 + w_{\text{DE}}|} \frac{1}{H_0}. \quad (16.7)$$

Using Eq. (2.36) with $h = 0.72$ for the equation of state $w_{\text{DE}} = -1.5$, we obtain $t_{\text{rip}} - t_0 \simeq 18 \text{ Gyr}$.

Chapter 3

- 3.1** Consider N points randomly distributed on the equatorial plane of a sphere of radius R . The density of points in the sphere is $n = N/V = 3N/(4\pi R^3)$. The superficial density on the plane is instead $n_s = N/(\pi R^2)$. The expected number of points inside a spherical shell of volume $4\pi r^2 dr$ will be equal to the area of the plane $2\pi r dr$ times the superficial density: $p(r) = 2\pi r n_s dr = 2N r dr/R^2$. To calculate the correlation function as in Eq. (3.5), we simply divide this by the expected number $p_P(r) = 4\pi r^2 n dr = 3N r^2 dr/R^3$ in a Poisson distribution of N points in the sphere (i.e. without confining the points to the plane). Then we have

$$\xi(r) = \frac{p(r)}{p_P(r)} - 1 = \frac{2R}{3r} - 1. \quad (16.8)$$

Therefore the correlation function $\eta(r) \equiv 1 + \xi(r)$ decreases as r^{-1} . This is the correlation function of a planar distribution in 3-dimensional space. This can be generalized to a d -dimensional distribution in a D -space to r^{d-D} . Conversely, a distribution whose correlation function $\eta(r)$ in D dimensions decreases as a power-law r^{d-D} is said to be d -dimensional. If d is not an integer, then the distribution is a fractal of dimension d . Notice that in this case the amplitude of the correlation $\eta(r)$ depends on R , i.e. on the “survey” size. This is a manifestation of the fact that a planar distribution in 3D is never close to homogeneity, no matter how large we take the survey volume. For the same reason, in this case the amplitude of the power spectrum depends on the survey size R and is totally unrelated to the level of homogeneity of the distribution.

3.2 For the Gaussian power spectrum given by $P = P_0 \exp[-k^2/(2\sigma_k^2)]$ the correlation function in Eq. (3.23) is

$$\xi(r) = \frac{P_0}{(2\pi)^3} \int \exp\left(-\frac{k^2}{2\sigma_k^2}\right) e^{i\mathbf{k}\cdot\mathbf{r}} d^3k. \quad (16.9)$$

We can integrate the angular part of $\mathbf{k}$:

$$\begin{aligned} \xi(r) &= \frac{P_0}{(2\pi)^3} \int_0^\infty k^2 \exp\left(-\frac{k^2}{2\sigma_k^2}\right) dk \int_0^\pi e^{ikr \cos \theta} \sin \theta d\theta \int_0^{2\pi} d\phi \\ &= \frac{P_0}{2\pi^2} \int_0^\infty k^2 \exp\left(-\frac{k^2}{2\sigma_k^2}\right) \frac{\sin(kr)}{kr} dk \\ &= \frac{P_0}{2r^3\pi^2} \int_0^\infty x \exp\left[-\frac{x^2}{2(r\sigma_k)^2}\right] \sin x dx \\ &= \sqrt{\frac{\pi}{2}} \frac{P_0\sigma_k^3}{2\pi^2} e^{-\frac{1}{2}\sigma_k^2 r^2}, \end{aligned} \quad (16.10)$$

where in the last line we used $\int_0^\infty x \exp[-x^2/(2c^2)] \sin x dx = \sqrt{\pi/2} c^3 e^{-c^2/2}$. The correlation function has a maximum at $r = 0$ and then declines to zero at large r . Although the power spectrum has a maximum at $k = 0$ (infinite wavelengths), the number of modes, proportional to $P(k)k^2 dk$, goes to zero for small k and therefore the distribution reaches homogeneity at large scales.

3.3 We proceed as in Eq. (3.53):

$$\begin{aligned} \langle N^3 \rangle &= \langle \sum n_i \sum n_j \sum n_k \rangle = \sum \langle n_i^3 \rangle + 3 \sum \langle n_i^2 \rangle \sum \langle n_i \rangle + \sum \langle n_i n_j n_k \rangle \\ &= N_0 + 3N_0^2 + N_0^3 \int W_i W_j W_k (1 + \xi_{ij} + \xi_{ik} + \xi_{jk} + \zeta_{ijk}) dV_i dV_j dV_k \\ &= N_0 + 3N_0^2 + N_0^3 + 3N_0^3 \int W_i W_j \xi_{ij} dV_i dV_j + N_0^3 \int W_i W_j W_k \zeta_{ijk} dV_i dV_j dV_k, \end{aligned}$$

where in the second line we used the definition of the three-point correlation given in Eq. (3.10) and in the third line we used the fact that the integrals over the functions ξ_{ij} are all identical. We then find the third-order moment

$$\begin{aligned} M_3 &= N_0^{-3} \langle (\Delta N)^3 \rangle = N_0^{-3} \langle N^3 - 3N^2 N_0 + 3N N_0^2 - N_0^3 \rangle \\ &= N_0^{-3} \langle N^3 \rangle - 3 \langle N^2 \rangle N_0^{-2} + 2. \end{aligned} \quad (16.11)$$

Now from Eq. (3.54) we have

$$\langle N^2 \rangle N_0^{-2} = N_0^{-1} + \int W_i W_j (1 + \xi_{ij}) dV_i dV_j = N_0^{-1} + 1 + \int W_i W_j \xi_{ij} dV_i dV_j, \quad (16.12)$$

so that finally

$$\begin{aligned}
 M_3 &= N_0^{-2} + 3N_0^{-1} + 3 \int W_i W_j \xi_{ij} dV_i dV_j + \int W_i W_j W_k \zeta_{ijk} dV_i dV_j dV_k \\
 &\quad - 3N_0^{-1} - 3 \int W_i W_j \xi_{ij} dV_i dV_j \\
 &= N_0^{-2} + \int W_i W_j W_k \zeta_{ijk} dV_i dV_j dV_k.
 \end{aligned} \tag{16.13}$$

Chapter 4

4.1 (1) If $P_{xy}(\delta_x, \delta_y)$ is the probability of having δ_x, δ_y in the interval $d\delta_x d\delta_y$, the probability of having both $\delta_x > \nu\sigma$ and $\delta_y > \nu\sigma$ is simply

$$\begin{aligned}
 P_2 &= \int_{>\nu\sigma}^{\infty} d\delta_x \int_{>\nu\sigma}^{\infty} d\delta_y P_{xy}(\delta_x, \delta_y) \\
 &= \frac{1}{2\pi(\sigma^4 - \xi_{12}^2)^{1/2}} \int_{>\nu\sigma}^{\infty} d\delta_x \int_{>\nu\sigma}^{\infty} d\delta_y e^{-\frac{1}{2(\sigma^4 - \xi_{12}^2)}(\sigma^2\delta_x^2 + \sigma^2\delta_y^2 - 2\xi_{12}\delta_x\delta_y)} \\
 &= \frac{1}{2\pi(\sigma^4 - \xi_{12}^2)^{1/2}} \int_{>\nu\sigma}^{\infty} d\delta_x e^{-\frac{(\sigma^2 - \xi_{12}^2\sigma^{-2})\delta_x^2}{2(\sigma^4 - \xi_{12}^2)}} \int_{>\nu\sigma}^{\infty} d\delta_y e^{-\frac{\sigma^2(\delta_y - \xi_{12}\delta_x\sigma^{-2})^2}{2(\sigma^4 - \xi_{12}^2)}} \\
 &= \frac{1}{2\sigma\sqrt{2\pi}} \int_{>\nu\sigma}^{\infty} d\delta_x e^{-\frac{\delta_x^2}{2\sigma^2}} \operatorname{erfc}\left(\frac{\nu\sigma - \xi_{12}\delta_x\sigma^{-2}}{\sqrt{2}(\sigma^2 - \sigma^{-2}\xi_{12}^2)^{1/2}}\right).
 \end{aligned} \tag{16.14}$$

The probability of having $\delta_x > \nu\sigma$ is similarly

$$P_1 = \int_{>\nu\sigma}^{\infty} d\delta_x P(\delta_x) = \frac{1}{2} \operatorname{erfc}\left(\frac{\nu}{\sqrt{2}}\right). \tag{16.15}$$

By the law of conditional probabilities the conditional probability of having $\delta_x > \nu\sigma$ given that $\delta_y > \nu\sigma$ is

$$P_c = \frac{P_2}{P_1}. \tag{16.16}$$

(2) If we divide a distribution into many small equal-volume regions above and below threshold $\nu\sigma$, then the fraction of regions above threshold at distance r from a region above threshold equals the conditional probability P_c we have just evaluated. If we consider a shell encompassing N regions at distance r of volume V_r around a region above threshold, then a number $N P_c$ will be above threshold and their numerical density will be $\rho_c = N P_c / V_r$. On the other hand the fraction of regions above threshold equals P_1 and by the same argument their numerical density will be $\rho_0 = N P_1 / V_r$. Therefore the requested ratio is P_2 / P_1^2 . Now this can be interpreted as in Eq. (3.5) as a conditional correlation function for regions above threshold:

$$1 + \xi_{>\nu\sigma} = \frac{\rho_c}{\rho_0} = \frac{P_2}{P_1^2}. \tag{16.17}$$

This amounts to

$$1 + \xi_{>\nu\sigma} = \frac{2 \int_{>\nu\sigma} d\delta_x e^{-\frac{\delta_x^2}{2\sigma^2}} \operatorname{erfc}\left(\frac{\sigma^3\nu - \xi_{12}\delta_x}{\sqrt{2}\sigma(\sigma^4 - \xi_{12}^2)^{1/2}}\right)}{\sigma\sqrt{2\pi}[\operatorname{erfc}(\nu/\sqrt{2})]^2}. \quad (16.18)$$

(3) In the limit of $\xi_{12} \ll 1$ we have

$$\operatorname{erfc}\left(\frac{\sigma^3\nu - \xi_{12}\delta_x}{\sqrt{2}\sigma(\sigma^4 - \xi_{12}^2)^{1/2}}\right) \simeq \operatorname{erfc}\left[x\left(1 - \frac{\xi_{12}\delta_x}{\nu\sigma^3}\right)\right] \simeq \operatorname{erfc}(x) + \frac{2}{\sqrt{\pi}} \frac{\xi_{12}\delta_x}{\nu\sigma^3} x e^{-x^2}, \quad (16.19)$$

where $x = \nu/\sqrt{2}$. We have then

$$\begin{aligned} 1 + \xi_{>\nu\sigma} &\simeq \frac{2 \int_{>\nu\sigma} d\delta_x e^{-\frac{\delta_x^2}{2\sigma^2}} \left(\operatorname{erfc}(x) + \frac{2}{\sqrt{\pi}} \frac{\xi_{12}\delta_x}{\nu\sigma^3} x e^{-x^2}\right)}{\sigma\sqrt{2\pi}[\operatorname{erfc}(x)]^2} \\ &\simeq 1 + \frac{4\xi_{12}x e^{-x^2}}{\nu\sigma^3\sqrt{\pi}[\operatorname{erfc}(x)]^2} \frac{1}{\sqrt{2\pi}\sigma} \int_{>\nu\sigma} d\delta_x \delta_x e^{-\frac{\delta_x^2}{2\sigma^2}} \\ &= 1 + \frac{4\xi_{12}x}{\nu\sigma^2\pi[\operatorname{erfc}(x)]^2} \frac{e^{-2x^2}}{\sqrt{2}}. \end{aligned} \quad (16.20)$$

Since $\operatorname{erfc}(x) \simeq e^{-x^2}/(x\sqrt{\pi})$ for large x , we finally have

$$\xi_{>\nu\sigma} \simeq \frac{4\xi_{12}x^3}{\sqrt{2}\nu\sigma^2} \simeq \left(\frac{\nu}{\sigma}\right)^2 \xi_{12}. \quad (16.21)$$

This shows that in the limit of $\xi_{12} \ll 1$ and $\nu \gg 1$ the correlation functions of regions above the threshold is $(\nu/\sigma)^2$ times the underlying correlation function. The ratio $b = \nu/\sigma$ can be seen then as the bias between the regions and the density field. This is an example of how a linear, constant bias can arise if we identify the regions above threshold as sites of galaxy formation.

4.2 Multiplying the term $\mathcal{P}_0(\mu) = 1$ for Eq. (4.170) and integrating it in the range $[-1, 1]$, we obtain

$$\begin{aligned} \Theta'_0 + \Phi' \int_{-1}^{+1} \frac{d\mu}{2} + ik \int_{-1}^{+1} \frac{d\mu}{2} \mu \Theta(\mu) + ik\Psi \int_{-1}^{+1} \frac{d\mu}{2} \mu \\ = -\tau'_{\text{op}} \left(\Theta_0 - \Theta_0 + v_b \int_{-1}^{+1} \frac{d\mu}{2} \mu \right), \end{aligned} \quad (16.22)$$

which shows that the r.h.s. of this equation vanishes. Since the dipole moment is given by $\Theta_1 = -(1/i) \int_{-1}^{+1} (d\mu/2) \mu \Theta(\mu)$, the above equation reduces to Eq. (4.176).

Similarly the multiplication of the term $\mathcal{P}_1(\mu) = \mu$ for Eq. (4.170) yields

$$\Theta'_1 - k \int_{-1}^{+1} \frac{d\mu}{2} \mu^2 \Theta(\mu) - \frac{1}{3} k \Psi = \tau'_{\text{op}} \left(\Theta_1 - \frac{i}{3} v_b \right). \quad (16.23)$$

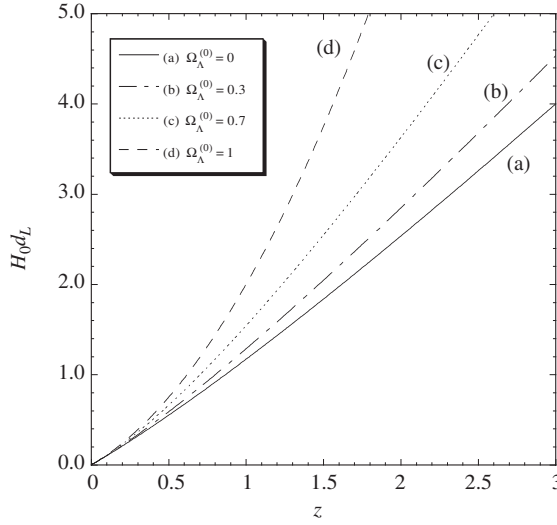


Figure 16.1 The luminosity distance d_L versus the redshift z in the flat Universe for $\Omega_\Lambda^{(0)} = 0, 0.3, 0.7, 1$.

The quadrupole moment is given by $\Theta_2 = \int_{-1}^{+1} (d\mu/2) \mathcal{P}_2(\mu) \Theta(\mu)$, where $\mathcal{P}_2(\mu) = (3\mu^2 - 1)/2$. It then follows that

$$\int_{-1}^{+1} \frac{d\mu}{2} \mu^2 \Theta(\mu) = \frac{1}{3} \Theta_0 - \frac{2}{3} \Theta_2. \quad (16.24)$$

Neglecting Θ_2 relative to Θ_0 and substituting Eq. (16.24) into Eq. (16.23), we obtain Eq. (4.177).

4.3 Substituting Eq. (4.178) into Eq. (4.177), we get

$$\Theta'_1 = \frac{1}{3} k \Psi + \frac{k}{3(1+R_s)} \Theta_0 - \frac{R_s}{1+R_s} \frac{a'}{a} \Theta_1. \quad (16.25)$$

Taking the derivative of Eq. (4.176) in terms of η and using Eq. (16.25) to eliminate Θ'_1 , we find

$$\Theta''_0 + \frac{1}{3} k^2 \Psi + \frac{k^2}{3(1+R_s)} \Theta_0 - \frac{R_s}{1+R_s} \frac{a'}{a} k \Theta_1 = -\Phi''. \quad (16.26)$$

After eliminating the term Θ_1 with the use of Eq. (4.177), we arrive at Eq. (4.179).

Chapter 5

5.1 In Fig. 16.1 we plot $d_L(z)$ versus z for several different values of $\Omega_\Lambda^{(0)}$ (derived by using a fortran code). The luminosity distance gets larger for increasing $\Omega_\Lambda^{(0)}$.

5.2 The numerical plot of $r_{\text{BAO}}(z)$ is given in Fig. 5.9.

Chapter 6

- 6.1** For the Kähler potential (6.36) and the superpotential $W(C)$ that depends on C only, the field potential is given by

$$V = \frac{\kappa^2}{(T + T^* - 2CC^*)^3(S + S^*)} \left[|D_C W|^2 \frac{(T + T^* - 2CC^*)^2}{6(T + T^*)} + |D_S W|^2 (S + S^*)^2 \right], \quad (16.27)$$

where

$$D_C W = \frac{\partial W}{\partial C} + \frac{6C^* W}{T + T^* - 2CC^*}, \quad D_S W = -\frac{W}{S + S^*}. \quad (16.28)$$

If the potential V vanishes we have that $D_C W = 0$ and $D_S W = 0$, thereby giving $W = 0$ and $\partial W / \partial C = 0$. In this case $D_T W = -3W / (T + T^*) = 0$ and hence the field configuration with $V = 0$ corresponds to the supersymmetric state.

- 6.2** For the Kähler potential (6.37) and the superpotential (6.40) with $\rho = i\sigma$, we obtain $D_\rho W = i [cAe^{-c\sigma} + (3/2)(W_0 + Ae^{-c\sigma})/\sigma]$ and $K_{\rho\rho^*} = 3/(4\sigma^2)$. We then find that the field potential V is given by Eq. (6.42). Taking the derivative of Eq. (6.42) in terms of σ , it follows that

$$\frac{dV}{d\sigma} = -\frac{cA}{6} e^{-2c\sigma} \sigma^{-3} (2 + c\sigma) [3W_0 e^{c\sigma} + A(3 + 2c\sigma)], \quad (16.29)$$

which has a minimum at the value $\sigma = \sigma_c$ satisfying the condition (6.41).

Chapter 7

- 7.1** The equation for perturbations about the fixed point is given by Eq. (7.25) with the matrix element

$$a_{11} = -3 + \frac{9}{2}(1 - w_M)x_1^2 + \frac{3}{2}(1 + w_M)(1 - x_2^2), \quad (16.30)$$

$$a_{12} = \sqrt{6}\lambda x_2 - 3(1 + w_M)x_1 x_2, \quad (16.31)$$

$$a_{21} = -\frac{\sqrt{6}}{2}\lambda x_2 + 3(1 - w_M)x_1 x_2, \quad (16.32)$$

$$a_{22} = -\frac{\sqrt{6}}{2}\lambda x_1 + \frac{3}{2}(1 - w_M)x_1^2 + \frac{3}{2}(1 + w_M)(1 - 3x_2^2). \quad (16.33)$$

From Eq. (7.26) we obtain the eigenvalues of this matrix for each fixed point.

- 7.2** Taking the derivative of the quantity $x = \dot{\phi}^2 / (2V)$ in terms of N , we find

$$1 + \frac{1}{6} \frac{d \ln x}{dN} = -\frac{V_{,\phi} \dot{\phi}^2 + 2V}{V \cdot 6H\dot{\phi}}. \quad (16.34)$$

Using the relation $\sqrt{(1 + w_\phi)/\Omega_\phi} = 2\sqrt{3}H|\dot{\phi}|/(\dot{\phi}^2 + 2V)$, we obtain Eq. (7.30). Differentiating Eq. (7.30) with respect to ϕ gives

$$\Gamma - 1 = \frac{V}{V_\phi} \frac{H}{\dot{\phi}} \left[\frac{w'_\phi}{2(1 + w_\phi)} - \frac{\Omega'_\phi}{2\Omega_\phi} + \frac{y''}{6 + y'} \right], \quad (16.35)$$

where a prime represents a derivative with respect to N . Substituting $w'_\phi/(1 + w_\phi) = y/(x + 1)$ and $\Omega'_\phi/\Omega_\phi = 3(1 - \Omega_\phi)[w_M(x + 1) - x + 1]/(x + 1)$ into Eq. (16.35), we arrive at Eq. (7.31).

7.3 We consider the two-field system in the presence of a barotropic fluid with an equation of state w_M . Defining the following dimensionless quantities

$$x_i \equiv \frac{\kappa \dot{\phi}_i}{\sqrt{6}H}, \quad y_i \equiv \frac{\kappa \sqrt{V_i e^{-\kappa \lambda_i \phi_i}}}{\sqrt{3}H}, \quad (16.36)$$

we obtain the following autonomous equations ($i = 1, 2$)

$$\frac{dx_i}{dN} = -3x_i + \lambda_i \sqrt{\frac{3}{2}} y_i^2 + \frac{3}{2} x_i [2x_1^2 + 2x_2^2 + (1 + w_M)(1 - x_1^2 - y_1^2 - x_2^2 - y_2^2)], \quad (16.37)$$

$$\frac{dy_i}{dN} = -\lambda_i \sqrt{\frac{3}{2}} x_i y_i + \frac{3}{2} y_i [2x_1^2 + 2x_2^2 + (1 + w_M)(1 - x_1^2 - y_1^2 - x_2^2 - y_2^2)], \quad (16.38)$$

together with

$$\frac{1}{H} \frac{dH}{dN} = -\frac{3}{2} [2x_1^2 + 2x_2^2 + (1 + w_M)(1 - x_1^2 - y_1^2 - x_2^2 - y_2^2)], \quad (16.39)$$

$$\Omega_M = 1 - x_1^2 - y_1^2 - x_2^2 - y_2^2. \quad (16.40)$$

Setting $dx_i/dN = dy_i/dN = 0$, we get eight fixed points for the above system [192]. Among them the stable fixed point that can be used for the late-time acceleration is $(x_1, y_1, x_2, y_2) = (\lambda_{\text{eff}}^2/\sqrt{6}\lambda_1, (\lambda_{\text{eff}}/\lambda_1)(1 - \lambda_{\text{eff}}^2/6)^{1/2}, \lambda_{\text{eff}}^2/\sqrt{6}\lambda_2, (\lambda_{\text{eff}}/\lambda_2)(1 - \lambda_{\text{eff}}^2/6)^{1/2})$ with $w_{\text{eff}} = -1 + \lambda_{\text{eff}}^2/3$ and $\Omega_\phi = 1$. Considering linear perturbations about this fixed point, we find that it is stable under the condition $\lambda_{\text{eff}}^2 < 3(1 + w_M)$.

7.4 Substituting the relations

$$K_{\varphi\varphi^*} = \frac{1 - \ln(2\kappa\varphi)}{2\kappa^2\varphi^2} = \frac{X}{2\kappa^2\varphi^2}, \quad D_i W = -M^{3+\alpha}\varphi^{-\alpha-1}(\alpha - 1 + X), \quad (16.41)$$

into Eq. (6.23), we get the potential (7.65).

7.5 Using the derivative with respect to z , Eq. (7.8) can be written as

$$\frac{\kappa^2}{2} \left(\frac{d\phi}{dz} \right)^2 = -\frac{1}{(1+z)^2} \left(\frac{\kappa^2 \rho_m}{2H^2} - \frac{1+z}{H} \frac{dH}{dz} \right). \quad (16.42)$$

This reduces to Eq. (7.85) by substituting $\rho_m = \rho_m^{(0)}(1+z)^3$. Plugging Eq. (7.85) into Eq. (7.7), we obtain Eq. (7.86).

Chapter 8

- 8.1** We introduce a new scalar field defined by $\varphi = e^{\kappa\alpha\lambda\phi}/(\kappa\alpha\lambda)$. Since $\partial\varphi/\partial\phi = e^{\kappa\alpha\lambda\phi}$ it follows that $X = \tilde{X}(\kappa\alpha\lambda\varphi)^{-2}$ and $Xe^{\kappa\lambda\phi} = \tilde{X}(\kappa\alpha\lambda\varphi)^{(1-2\alpha)/\alpha}$, where $\tilde{X} \equiv -(1/2)g^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi$. Hence the scaling Lagrangian (8.27) can be written as

$$P(\varphi, X) = (\kappa\alpha\lambda\varphi)^{-2}\tilde{X}g(\tilde{X}(\kappa\alpha\lambda\varphi)^{(1-2\alpha)/\alpha}). \quad (16.43)$$

For the choice $\alpha = 1/2$ this Lagrangian density reduces to $P(\varphi, \tilde{X}) = K(\varphi)p(\tilde{X})$ with

$$K(\varphi) = (\kappa\lambda/2)^{-2}\varphi^{-2}, \quad p(\tilde{X}) = \tilde{X}g(\tilde{X}). \quad (16.44)$$

This shows that the k-essence model with the Lagrangian density $P(\phi, X) = V_0\phi^{-2}p(X)$ has a scaling solution.

- 8.2** Substituting the relations $P_\phi = K(\phi)g(y)/y$ and $\rho_\phi = -K(\phi)g'(y)$ into Eq. (8.15), we obtain

$$\frac{dy}{dN} = \frac{3}{g''(y)} \left[\frac{g(y) - yg'(y)}{y} - \frac{\dot{K}}{3HK}g'(y) \right]. \quad (16.45)$$

For the choice $K(\phi) = 1/\phi^2$ one has

$$\frac{\dot{K}}{HK} = -\frac{2\dot{\phi}}{H\phi} = -\frac{2\sqrt{6}}{y} \sqrt{\frac{\Omega_\phi}{-g'(y)}}, \quad (16.46)$$

where we have used the relations $1/(H\phi) = \sqrt{3\Omega_\phi/(-g'(y))}$ and $\dot{\phi} = \sqrt{2}/y$. Substituting Eq. (16.46) into Eq. (16.45), we arrive at Eq. (8.53).

Taking the derivative of the relation $\Omega_\phi = \rho_\phi/(3H^2)$ in terms of N , it follows that

$$\frac{d\Omega_\phi}{dN} = \Omega_\phi \left[-3(1 + w_\phi) - 2\frac{\dot{H}}{H^2} \right], \quad (16.47)$$

where we have used $\dot{\rho}_\phi/(H\rho_\phi) = -3(1 + w_\phi)$. From Eq. (8.14) we find

$$\begin{aligned} 2\frac{\dot{H}}{H^2} &= -\frac{1}{H^2\phi^2} \frac{g(y) - yg'(y)}{y} - \frac{3\rho_m + 4\rho_r}{3H^2} \\ &= -3\Omega_\phi(1 + w_\phi) - 3(1 - \Omega_\phi)(1 + w_{rm}). \end{aligned} \quad (16.48)$$

At the second equality we have employed the relations $g(y) - yg'(y) = -yg'(y)(1 + w_\phi)$ and $(3\rho_m + 4\rho_r)/(3H^2) = 3(1 - \Omega_\phi)(1 + w_m)$. Substituting Eq. (16.48) into Eq. (16.47) results in Eq. (8.54).

Using the definition of w_{rm} and the relation $\dot{\rho}_r = -4H\rho_r$, we can obtain Eq. (8.55) easily.

8.3 Perturbing Eqs. (8.82)–(8.84) about the fixed points (x_1, x_2, x_3) , we obtain the perturbation equations

$$\begin{aligned} \frac{d}{dN}\delta x_1 = & \left(-\frac{3}{2} + \frac{9}{2}x_1^2 + \sqrt{6}Qx_1 - \frac{3}{2}x_2^2 + \frac{1}{2}x_3^2\right)\delta x_1 \\ & + x_2 \left[\sqrt{6}(Q + \lambda) - 3x_1\right]\delta x_2 + x_3(x_1 + \sqrt{6}Q)\delta x_3, \end{aligned} \quad (16.49)$$

$$\frac{d}{dN}\delta x_2 = x_2 \left(3x_1 - \frac{\sqrt{6}}{2}\lambda\right)\delta x_1 + \frac{1}{2}(3 + 3x_1^2 - 9x_2^2 + x_3^2 - \sqrt{6}\lambda x_1)\delta x_2 + x_2 x_3 \delta x_3, \quad (16.50)$$

$$\frac{d}{dN}\delta x_3 = 3x_1 x_3 \delta x_1 - 3x_2 x_3 \delta x_2 + \frac{1}{2}(-1 + 3x_1^2 - 3x_2^2 + 3x_3^2)\delta x_3. \quad (16.51)$$

For the fixed points (e), (a), (c) the eigenvalues of the 3×3 Jacobian matrix of perturbations are given by Eqs. (8.85), (8.86), and (8.88), respectively.

8.4 At the linear level the perturbation of the quantity Y about Y_c is given by $\delta Y = 2[(x_c/y_c^2)\delta x - (x_c^2/y_c^3)\delta y]$. When we consider linear perturbations of the term $g + g_1$, we need to take into account the second-order derivative of g in terms of Y . Then it follows that $\delta(g + Yg') = (2g'_c + Y_c g''_c)\delta Y$, where $g'_c \equiv dg(Y_c)/dY$. The perturbation of the density parameter, $\Omega_\varphi = x^2(g + 2Yg')$, is

$$\delta\Omega_\varphi = 2(x_c/A_c)\delta x - 2(x_c^2/y_c)(3Y_c g'_c + 2Y_c^2 g''_c)\delta y, \quad (16.52)$$

where $A_c = [g_c + 5Y_c g'_c + 2Y_c^2 g''_c]^{-1}$. Recalling that the fixed points A and B satisfy Eqs. (8.210) and (8.211), we find that the components of the Jacobian matrix $\mathcal{M}$ of perturbations are

$$a_{11} = -3 + \frac{\sqrt{6}}{2}(2Q + \lambda)x_c + 3x_c^2(g_c + Y_c g'_c), \quad (16.53)$$

$$a_{12} = y_c \left[-3g'_c x_c Y_c^2 + \frac{3x_c}{y_c^2} - \sqrt{6}(Q + \lambda)Y_c + \sqrt{6}A_c \frac{(Q + \lambda)\Omega_\varphi + Q}{2y_c^2} \right], \quad (16.54)$$

$$a_{21} = \frac{y_c}{2} \left[-\sqrt{6}\lambda + 6x_c(g_c + Y_c g'_c) \right], \quad (16.55)$$

$$a_{22} = -3g'_c x_c^2 Y_c. \quad (16.56)$$

We then obtain the eigenvalues (8.213) and (8.219) for the fixed points A and B, respectively.

Chapter 9

9.1 The variation of the action (9.226) with respect to $g_{\mu\nu}$ and ϕ leads to the following equations

$$F G_{\mu\nu} = \frac{1}{2}(f - RF)g_{\mu\nu} + \nabla_\nu F_{,\mu} - g_{\mu\nu} \square F + \frac{1}{2}f_{,X}\phi_{,\mu}\phi_{,\nu} + T_{\mu\nu}^{(m)}, \quad (16.57)$$

$$\nabla_c(f_{,X}\phi^{,c}) + f_{,\phi} = 0, \quad (16.58)$$

where $F = \partial f / \partial R$ and $T_{\mu\nu}^{(m)}$ is the energy-momentum tensor of non-relativistic matter. In the flat FLRW spacetime Eqs. (16.57) and (16.58) give

$$3FH^2 = f_{,X}X + \frac{1}{2}(FR - f) - 3H\dot{F} + \rho_m, \quad (16.59)$$

$$-2F\dot{H} = f_{,X}X + \ddot{F} - H\dot{F} + \rho_m, \quad (16.60)$$

$$\frac{1}{a^3} (a^3 \dot{\phi} f_{,X})' - f_{,\phi} = 0. \quad (16.61)$$

We define

$$\rho_{\text{DE}} \equiv f_{,X}X + \frac{1}{2}(FR - f) - 3H\dot{F} + 3H^2(A - F), \quad (16.62)$$

$$P_{\text{DE}} \equiv \ddot{F} + 2H\dot{F} - \frac{1}{2}(FR - f) - (2\dot{H} + 3H^2)(A - F), \quad (16.63)$$

where A is some constant. Then Eqs. (16.59) and (16.60) can be written as Eqs. (9.44) and (9.45) with $\rho_r = 0$. Moreover one can easily show that ρ_{DE} and P_{DE} defined above satisfy the continuity equation $\dot{\rho}_{\text{DE}} + 3H(\rho_{\text{DE}} + P_{\text{DE}}) = 0$.

9.2 Expanding Eq. (9.124) under the linear expansion of the variable $\psi = \psi_0(1 + \delta_\psi)$, we find

$$\left(\frac{\partial^2}{\partial t^2} - \nabla^2 \right) \delta_\psi + M_\psi^2 \delta_\psi = - \frac{\delta T}{(3 + 2\omega_{\text{BD}})\psi_0}, \quad (16.64)$$

where

$$M_\psi^2 \equiv \frac{2(\psi_0 U_{,\psi\psi} - U_{,\psi})}{3 + 2\omega_{\text{BD}}}. \quad (16.65)$$

Since $\psi = F(R)$, $U = (RF - f)/2$, and $\omega_{\text{BD}} = 0$ for $f(R)$ theory in the metric formalism, we have $U_{,\psi} = R/2$ and $U_{,\psi\psi} = 1/(2f_{,RR})$. In this case we obtain the field mass squared (9.52) from Eq. (16.65).

9.3 The eigenvalues for the fixed points (a)–(e) are given by

$$(a) -\frac{3-2Q^2}{2(1-2Q^2)}, \quad \frac{3+2Q\lambda-6Q^2}{2(1-2Q^2)}, \quad (16.66)$$

$$(b1) \frac{3(\sqrt{6}+4Q-\lambda)}{\sqrt{6}+6Q}, \quad \frac{3+\sqrt{6}Q}{1+\sqrt{6}Q}, \quad (16.67)$$

$$(b2) \frac{3(\sqrt{6}-4Q+\lambda)}{\sqrt{6}-6Q}, \quad \frac{3-\sqrt{6}Q}{1-\sqrt{6}Q}, \quad (16.68)$$

$$(c) -\frac{6-\lambda^2+8Q\lambda-16Q^2}{2(1-4Q^2+Q\lambda)}, \quad -\frac{3-\lambda^2+7Q\lambda-12Q^2}{1-4Q^2+Q\lambda}, \quad (16.69)$$

$$(d) \frac{3(2Q-\lambda)}{4\lambda} \left[1 \pm \sqrt{1 + \frac{8(6Q^2-2Q\lambda-3)(12Q^2+\lambda^2-7Q\lambda-3)}{3(2Q-\lambda)^2}} \right], \quad (16.70)$$

$$(e) -3, \quad -3. \quad (16.71)$$

9.4 Taking the derivative of Eq. (9.212) in terms of τ , we obtain

$$\frac{[\dot{a}']}{a_b} = -\frac{\kappa_{(5)}^2}{3} \dot{\rho}_M - \frac{\kappa_{(5)}^2}{3} \frac{\dot{a}_b}{a_b} \rho_M - \frac{\dot{a}_b}{a_b} \frac{\kappa_{(5)}^2}{\kappa_{(4)}^2 n_b^2} \left(\frac{\dot{a}_b^2}{a_b^2} + 2 \frac{\dot{a}_b}{a_b} \frac{\dot{n}_b}{n_b} - 2 \frac{\ddot{a}_b}{a_b} + K \frac{n_b^2}{a_b^2} \right). \quad (16.72)$$

Substituting Eqs. (16.72) and (9.213) into the equation, $(\dot{a}_b/a_b)([n']/n_b) - [\dot{a}']/a_b = 0$, we get

$$\frac{d\rho_M}{d\tau} + 3 \frac{\dot{a}_b}{a_b} (\rho_M + P_M) = 0, \quad (16.73)$$

which is equivalent to Eq. (9.217).

Chapter 10

10.1 The component R_{01} of the Ricci tensor for the metric equation (10.1) is

$$R_{01} = \frac{2R'\dot{X} - 2X\dot{R}'}{XR}. \quad (16.74)$$

Therefore $R_{01} = 0$ implies

$$\frac{\dot{X}}{X} = \frac{\dot{R}'}{R'}, \quad (16.75)$$

which is solved by

$$X(r, t) = R'(r, t)f(r), \quad (16.76)$$

where $f(r)$ is an arbitrary function of r . By choosing $\beta(r) = f^{-2} - 1$ we recover the LTB metric in the form of Eq. (10.2).

10.2 Let us rewrite Eq. (10.10) as

$$H_{\perp}^2 = \left(\frac{\dot{R}}{R} \right)^2 = H_{\perp,0}^2 \left[\frac{\Omega_m^{(0)} R_0^3}{R^3} + \frac{\Omega_K^{(0)} R_0^2}{R^2} \right], \quad (16.77)$$

where $\Omega_K^{(0)} = 1 - \Omega_m^{(0)}$. Using the relation $dt = dR/(H_{\perp} R) = dx/(H_{\perp} x)$, where $x = R/R_0$, the cosmic age t_{age} is given by

$$H_{\perp,0} t_{\text{age}} = \int_0^1 \frac{dx}{[\Omega_m^{(0)}/x + \Omega_K^{(0)}]^{1/2}} = \frac{1}{\Omega_K^{(0)}} - \frac{\Omega_m^{(0)}}{(\Omega_K^{(0)})^{3/2}} \sinh^{-1} \sqrt{\frac{\Omega_K^{(0)}}{\Omega_m^{(0)}}}. \quad (16.78)$$

Therefore, if one chooses the Hubble function and the $\Omega_m^{(0)}$ function so that

$$H_{\perp,0} = T_0^{-1} \left[\frac{1}{\Omega_K^{(0)}} - \frac{\Omega_m^{(0)}}{(\Omega_K^{(0)})^{3/2}} \sinh^{-1} \sqrt{\frac{\Omega_K^{(0)}}{\Omega_m^{(0)}}} \right], \quad (16.79)$$

then the time since big bang, $t_{\text{age}} = T_0$, would be the same for every observer. (From Ref. [452].)

10.3 From the redshift

$$z \equiv \frac{\varepsilon(0) - \varepsilon(\lambda_s)}{\varepsilon(\lambda_s)}, \quad (16.80)$$

we obtain the derivative

$$\frac{dz}{d\lambda_s} = -\frac{d\varepsilon}{d\lambda_s} \frac{\varepsilon(0)}{\varepsilon^2(\lambda_s)} = -(1+z) \frac{d\varepsilon}{d\lambda_s} \frac{1}{\varepsilon}. \quad (16.81)$$

The next step is to derive $d\varepsilon/d\lambda_s$. Let us rewrite the $dt/d\lambda_s$ geodesic equation for $J = 0$ for a light ray $t_1(\lambda_s)$ [see Eq. (10.23)]:

$$\frac{dt_1}{d\lambda_s} = -\frac{dr}{d\lambda_s} \frac{R'(t_1, r)}{\sqrt{1 + \beta(r)}}. \quad (16.82)$$

Consider now the geodesic of the same light ray: $t_2(\lambda_s) = t_1(\lambda_s) + \varepsilon(\lambda_s)$. At first-order in ε we have

$$\frac{dt_2}{d\lambda_s} = \frac{dt_1}{d\lambda_s} + \frac{d\varepsilon}{d\lambda_s} = -\frac{dr}{d\lambda_s} \frac{R'(t_1 + \varepsilon, r)}{\sqrt{1 + \beta(r)}} = -\frac{dr}{d\lambda_s} \frac{R'(t_1) + \varepsilon \dot{R}'(t_1, r)}{\sqrt{1 + \beta(r)}}. \quad (16.83)$$

By subtraction we obtain

$$\frac{d\varepsilon}{d\lambda_s} = -\frac{dr}{d\lambda_s} \frac{\varepsilon \dot{R}'(t_1, r)}{\sqrt{1 + \beta(r)}}. \quad (16.84)$$

The redshift is then found as

$$\frac{dz}{d\lambda_s} = \frac{dr}{d\lambda_s} \frac{(1+z) \dot{R}'(t, r)}{\sqrt{1 + \beta(r)}}, \quad (16.85)$$

or

$$\frac{dz}{dr} = \frac{(1+z)\dot{R}'(t,r)}{\sqrt{1+\beta(r)}}, \quad (16.86)$$

which corresponds to Eq. (10.24). (From Ref. [453].)

Chapter 11

11.1 The equation to be solved is

$$\delta_m'' + \frac{1}{2}(1 - 3\Omega_{\text{DE}}w_{\text{DE}})\delta_m' - \frac{3}{2}\Omega_m\delta_m = 0. \quad (16.87)$$

If both $w_{\text{eff}} = \Omega_{\text{DE}}w_{\text{DE}}$ and $\Omega_m = 1 - \Omega_{\text{DE}}$ are constants, the above equation can be solved immediately by substituting $\delta_m = Ae^{\alpha N}$ where $N = \ln a$ is the number of e-foldings. We then obtain

$$\alpha^2 + \frac{1}{2}(1 - 3w_{\text{eff}})\alpha - \frac{3}{2}\Omega_m = 0, \quad (16.88)$$

which gives

$$\alpha = \frac{1}{4} \left[-1 + 3w_{\text{eff}} \pm \sqrt{24\Omega_m + (1 - 3w_{\text{eff}})^2} \right]. \quad (16.89)$$

11.2 As long as the oscillating mode can be neglected relative to the matter-induced mode in Eq. (11.136), we obtain $\delta F \simeq \delta\rho_m/[3(k^2/a^2 + M^2)]$. Plugging this into Eq. (11.135), we find

$$\Psi \simeq -\frac{\delta\rho_m}{2F} \frac{a^2}{k^2} \frac{4k^2/a^2 + 3M^2}{3(k^2/a^2 + M^2)}. \quad (16.90)$$

Then the matter perturbation equation (11.131) reduces to

$$\delta_m'' + \left(1 + \frac{\mathcal{H}'}{\mathcal{H}}\right)\delta_m' - \frac{3}{2}\Omega_m\delta_m \frac{4k^2/a^2 + 3M^2}{3(k^2/a^2 + M^2)} \simeq 0. \quad (16.91)$$

This gives the effective gravitational “constant”:

$$G_{\text{eff}} = \frac{G}{F} \frac{4k^2/a^2 + 3M^2}{3(k^2/a^2 + M^2)}, \quad (16.92)$$

where we recovered the gravitational constant G .

Chapter 12

12.1 Substituting the relations $\rho = \rho_m + \delta\rho_m$ and $\mathbf{u} = \dot{a}\mathbf{x} + \mathbf{v}$ into Eq. (12.1) and taking the zero-th-order part, we find

$$\dot{\rho}_m + \frac{1}{a}\nabla_x(\dot{a}\rho_m\mathbf{x}) = 0 \quad \rightarrow \quad \dot{\rho}_m + 3H\rho_m = 0. \quad (16.93)$$

The perturbed part corresponds to the equation

$$\frac{\partial(\delta\rho_m)}{\partial t} - H(\mathbf{x} \cdot \nabla_x)\delta\rho_m + \frac{1}{a}\rho_m\nabla_x\mathbf{v} + H\nabla_x \cdot (\delta\rho_m\mathbf{x}) + \frac{1}{a}\nabla_x \cdot (\delta\rho_m\mathbf{v}) = 0. \quad (16.94)$$

On using Eq. (16.93) and the relation $H\nabla_x \cdot (\delta\rho_m\mathbf{x}) = H(\mathbf{x} \cdot \nabla_x)\delta\rho_m + 3H\delta\rho_m$, Eq. (16.94) reduces to Eq. (12.7).

On using Eqs. (12.4) and (12.5), the following relations hold

$$\left(\frac{\partial\mathbf{u}}{\partial t}\right)_r = \ddot{a}\mathbf{x} + \left(\frac{\partial\mathbf{v}}{\partial t}\right)_x - H\dot{a}\mathbf{x} - H(\mathbf{x} \cdot \nabla_x)\mathbf{v}, \quad (16.95)$$

$$(\mathbf{u} \cdot \nabla_r)\mathbf{u} = H\dot{a}\mathbf{x} + H\mathbf{v} + H(\mathbf{x} \cdot \nabla_x)\mathbf{v} + \frac{1}{a}(\mathbf{v} \cdot \nabla_x)\mathbf{v}. \quad (16.96)$$

Substituting these relations into Eq. (12.2), we obtain Eq. (12.8) with Φ given in Eq. (12.10).

Equation (12.3) can be written as

$$\nabla_x^2 \left(\Phi_N - \frac{2}{3}\pi G a^2 \rho_m \mathbf{x}^2 \right) = 4\pi G a^2 \delta\rho_m. \quad (16.97)$$

On using the relation $\ddot{a}/a = -(4\pi G/3)\rho_m$, we obtain Eq. (12.9).

12.2 At second order Eq. (12.21) can be written as

$$\begin{aligned} \delta^{(2)''} + \left(1 + \frac{\mathcal{H}'}{\mathcal{H}}\right)\delta^{(2)'} - \frac{3}{2}\Omega_m\delta^{(2)} \\ = \delta_1^2 D_L^2 \left(\frac{3}{2}\frac{\Omega_m}{f^2} + 1\right) + (\nabla^i \delta_1) \left(\frac{v_i'}{\mathcal{H}} D_L + 2\frac{v_i}{\mathcal{H}} D_L' + \frac{v_i}{\mathcal{H}} D_L\right) + \frac{(\nabla^i v_j)(\nabla^j v_i)}{\mathcal{H}^2}. \end{aligned} \quad (16.98)$$

In the absence of shear the linear velocity is irrotational, so we can define a velocity potential v_p such that $v_i \equiv \nabla_i v_p$. Then Eq. (12.11) is of the Poisson type at first order and it can be solved to give

$$v_p = \mathcal{H} D_L' \Delta. \quad (16.99)$$

Then finally we can write Eq. (12.12) at first order as

$$v_i' = \left(-1 + \frac{3}{2}\frac{\Omega_m}{f}\right) \mathcal{H} D_L' \nabla_i \Delta, \quad (16.100)$$

and use it to simplify Eq. (16.98), which now becomes as in Eq. (12.22).

12.3 The virialization condition gives

$$R_V = R_T/2, \quad (16.101)$$

where R_T is the turnaround radius. The time t_V at virialization is twice the time t_T it takes to turnaround. Therefore the scale factor that grows as $a \propto t^{2/3}$ in the Einstein–de Sitter Universe is $a_V = 2^{2/3} a_T$ at virialization. The density contrast inside the

radius R is defined as

$$\delta = (aR_0/R)^3 - 1. \quad (16.102)$$

Then at R_V one has $\delta_V = (a_V R_0/R_V)^3 - 1 \simeq 32(a_T R_0/R_T)^3 - 1$ while at turnaround $\delta_T = (a_T R_0/R_T)^3 - 1$, so that we obtain

$$\delta_V \simeq 32(1 + \delta_T) - 1. \quad (16.103)$$

Since we have already calculated in Section 12.3 that $\delta_T \simeq 4.6$, we obtain $\delta_V \simeq 178$. The value of τ corresponding to this is $\simeq 4.77$, which in turn implies that $\delta_L \simeq 1.59$. This is very close to the singular value $\delta_c \simeq 1.686$, as expected.

12.4 The expression

$$\rho_{DE} \sim R^{-3(1-\beta)(1+w)} a^{-3\beta(1+w)} \quad (16.104)$$

is the solution of the conservation equation

$$\dot{\rho}_{DE} + 3\frac{\dot{R}}{R}(1+w)\rho_{DE} = \beta\Gamma\rho_{DE} \quad (16.105)$$

where $\Gamma = 3(1+w)(\dot{R}/R - \dot{a}/a)$. (From Refs. [520, 521].)

Chapter 13

13.1 We have

- the probability that a hypothesis is correct: $P(C) = 0.01$,
- the probability that a hypothesis is in agreement with data given that it is correct: $P(A|C) = 0.8$,
- the probability that a hypothesis is in agreement with data although it is wrong: $P(A|\bar{C}) = 0.1$.

From the first we obtain $P(\bar{C}) = 0.99$. From Bayes' theorem the probability for a hypothesis to be correct given that it is confirmed by data is

$$P(C|A) = \frac{P(A|C)P(C)}{P(A)}. \quad (16.106)$$

We can evaluate $P(A)$ using the definition of conditional probabilities as follows:

$$\begin{aligned} P(A) &= P(A, C) + P(A, \bar{C}) = P(A|C)P(C) + P(A|\bar{C})P(\bar{C}) \\ &= 0.8 \cdot 0.01 + 0.1 \cdot 0.99 \approx 0.107. \end{aligned} \quad (16.107)$$

Then we have

$$P(C|A) = \frac{P(A|C)P(C)}{P(A)} = \frac{0.8 \cdot 0.01}{0.107} \simeq 0.075. \quad (16.108)$$

That is, given our assumptions, a hypothesis confirmed by data is correct less than 8% of the times.

13.2 For the likelihood L_1 we need to perform the integral (we ignore the normalization constant)

$$\begin{aligned} L &= \int_{-\infty}^{\infty} d\alpha \exp \left[-\frac{1}{2} \sum_i \frac{(\alpha m_i - \mu_i)^2}{\sigma_i^2} \right] \\ &= e^{-\frac{1}{2} \sum_i \frac{\mu_i^2}{\sigma_i^2}} \int_{-\infty}^{\infty} d\alpha \exp \left[-\frac{1}{2} \sum_i \frac{\alpha^2 m_i^2 - 2\alpha \mu_i m_i}{\sigma_i^2} \right]. \end{aligned} \quad (16.109)$$

We define now

$$S_{ab} \equiv \sum_i \frac{m_i^a \mu_i^b}{\sigma_i^2}, \quad (16.110)$$

and complete the square in the integrand exponential by summing and subtracting S_{11}^2/S_{20} as follows

$$\begin{aligned} L &= e^{-\frac{1}{2} \left(S_{02} - \frac{S_{11}^2}{S_{20}} \right)} \int_{-\infty}^{\infty} d\alpha \exp \left[-\frac{1}{2} \left(\alpha^2 S_{20} - 2\alpha S_{11} + \frac{S_{11}^2}{S_{20}} \right) \right] \\ &= e^{-\frac{1}{2} \left(S_{02} - \frac{S_{11}^2}{S_{20}} \right)} \int_{-\infty}^{\infty} d\alpha \exp \left[-\frac{1}{2} \left(\alpha \sqrt{S_{20}} - \frac{S_{11}}{\sqrt{S_{20}}} \right)^2 \right]. \end{aligned} \quad (16.111)$$

The Gaussian integral gives $\sqrt{2\pi/S_{20}}$, which does not depend on the theoretical parameters and can be absorbed in the normalization constant N . Therefore the marginalized function becomes

$$L = N e^{-\frac{1}{2} \left(S_{02} - \frac{S_{11}^2}{S_{20}} \right)}. \quad (16.112)$$

For the likelihood L_2 we need to perform the integral

$$\begin{aligned} L &= \int_{-\infty}^{\infty} d\alpha \exp \left[-\frac{1}{2} \sum_i \frac{(m_i - \alpha \mu_i)^2}{\sigma_i^2} \right] \\ &= e^{-\frac{1}{2} \sum_i \frac{m_i^2}{\sigma_i^2}} \int_{-\infty}^{\infty} d\alpha \exp \left[-\frac{1}{2} \sum_i \frac{\alpha^2 \mu_i^2 - 2\alpha \mu_i m_i}{\sigma_i^2} \right]. \end{aligned} \quad (16.113)$$

We complete the square as follows

$$\begin{aligned} L &= e^{-\frac{1}{2} \left(S_{20} - \frac{S_{11}^2}{S_{02}} \right)} \int_{-\infty}^{\infty} d\alpha \exp \left[-\frac{1}{2} \left(\alpha^2 S_{02} - 2\alpha S_{11} + \frac{S_{11}^2}{S_{02}} \right) \right] \\ &= e^{-\frac{1}{2} \left(S_{20} - \frac{S_{11}^2}{S_{02}} \right)} \int_{-\infty}^{\infty} d\alpha \exp \left[-\frac{1}{2} \left(\alpha \sqrt{S_{02}} - \frac{S_{11}}{\sqrt{S_{02}}} \right)^2 \right]. \end{aligned} \quad (16.114)$$

The Gaussian integral gives $\sqrt{2\pi/S_{02}}$, which now depends on the theoretical parameters and must be included in the final likelihood. On the other hand the S_{20} term in

the first factor can be absorbed in the normalization. So we have

$$L = Ne^{\frac{1}{2}(\frac{S_{11}^2}{S_{02}} - \ln S_{02})}. \quad (16.115)$$

- 13.3** The prior of model A is a Dirac δ_D function centered on $\theta = 0$, whereas the prior of model B is $e^{-\theta^2/2\Sigma^2}/\sqrt{2\pi\Sigma^2}$. The data are described by a Gaussian $e^{-(\theta-\theta_{\max})/2\sigma^2}$ with $\theta_{\max} = \lambda\sigma$. We calculate Bayes' ratio as

$$\begin{aligned} B_{AB} &= \frac{\int f(\mathbf{x}; \theta_i^{M_1}) p(\theta_i^{M_1}) d\theta_i^{M_1}}{\int f(\mathbf{x}; \theta_i^{M_2}) p(\theta_i^{M_2}) d\theta_i^{M_2}} \\ &= \frac{\int e^{-(\theta-\theta_{\max})/2\sigma^2} \delta(\theta) d\theta}{(2\pi\Sigma^2)^{-1/2} \int e^{-(\theta-\theta_{\max})/2\sigma^2} e^{-\theta^2/2\Sigma^2} d\theta} \\ &= \sqrt{1+r^{-2}} e^{-\frac{\lambda^2}{2(1+r^2)}}, \end{aligned} \quad (16.116)$$

where $r = \sigma/\Sigma$. If the best-fit parameter $\theta_{\max}$ is many σ away from the predicted $\theta = 0$ (i.e. $\lambda \gg 1$), then it follows that $B_{AB} \ll 1$, favoring model B that allows for the extra freedom Σ . But if λ is not too large and $r \ll 1$, i.e. the data is much more peaked than the B prior and close to the predicted value, then we have $B_{AB} \approx 1/r \gg 1$ so that the extra parameter introduced by model B is not needed and A is favored. This is in line with Ockham's razor argument. If $r \gg 1$, then $B_{AB} \approx 1$ and hence there is not enough information to decide between A and B. Although B has more parameters, the fact that the data have a large error and are too poor to constrain θ implies that no preference must be given to either A or B.

Chapter 14

- 14.1** We need to evaluate

$$C_\ell = \frac{9}{2} \left(\frac{\Gamma[(\ell+1)/2]}{\Gamma[(\ell+2)/2]} \right)^2 \int \frac{dk}{k^4} P_\delta(k) [(\mathcal{H}^2 D)_{,\eta_p}]^2, \quad (16.117)$$

where η_p is the solution of

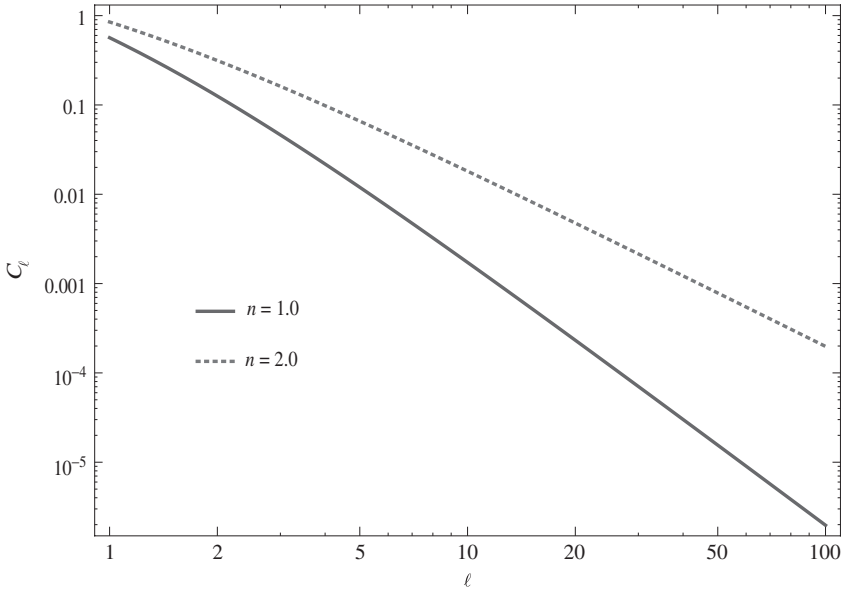
$$\chi(\eta_p) = \frac{\ell + 1/2}{k}. \quad (16.118)$$

Since $w_{\text{eff}} = \text{constant}$ we have $H^2 = H_0^2(1+z)^{3(1+w_{\text{eff}})}$ and

$$\chi(z_p) = \int_0^{z_p} \frac{d\tilde{z}}{H(\tilde{z})} = \frac{2}{H_0} \frac{1 - (1+z_p)^{-(1+3w_{\text{eff}})/2}}{1+3w_{\text{eff}}}, \quad (16.119)$$

from which Eq. (16.118) gives

$$1 + z_p = \left[1 - \frac{H_0(2\ell+1)(1+3w_{\text{eff}})}{4k} \right]^{-2/(1+3w_{\text{eff}})} = \left(1 - \frac{k_0}{k} \right)^{-2/(1+3w_{\text{eff}})}, \quad (16.120)$$

Figure 16.2 ISW spectra for $n = 1, 2$ in arbitrary units.

where $k_0 = H_0(2\ell + 1)(1 + 3w_{\text{eff}})/4$. Moreover, we have

$$\mathcal{H}^2 D = H^2 a^2 D = H_0^2 a^{p-1-3w_{\text{eff}}}, \quad (16.121)$$

$$(\mathcal{H}^2 D)_{,\eta} = \mathcal{H} \frac{d(\mathcal{H}^2 D)}{dN} = (p - 1 - 3w_{\text{eff}}) \mathcal{H}^3 D = (p - 1 - 3w_{\text{eff}}) H_0^3 a^{p-\frac{3}{2}(1+3w_{\text{eff}})}, \quad (16.122)$$

so that for $n < -3$ we have

$$\begin{aligned} C_\ell &= \frac{9}{2} \left(\frac{\Gamma[(\ell+1)/2]}{\Gamma[(\ell+2)/2]} \right)^2 A H_0^6 (1 + 3w_{\text{eff}} - p)^2 \int_0^\infty dk k^{n-4} (1 + z_p)^{3(1+3w_{\text{eff}})-2p} \\ &= A_1 \left(\frac{\Gamma[(\ell+1)/2]}{\Gamma[(\ell+2)/2]} \right)^2 k_0^{n-3} \int_0^\infty dq q^{n-4} \left(1 - \frac{1}{q} \right)^{-\frac{2(3-2p+9w_{\text{eff}})}{1+3w_{\text{eff}}}} \\ &= A_2 \left(\frac{\Gamma[(\ell+1)/2]}{\Gamma[(\ell+2)/2]} \right)^2 \left(\ell + \frac{1}{2} \right)^{n-3}, \end{aligned} \quad (16.123)$$

where $q = k/k_0$, $A_1 = 9AH_0^6(1 + 3w_{\text{eff}} - p)^2/2$, and $A_2 = A_1(H_0|1 + 3w_{\text{eff}}|/2)^{n-3} \Gamma[3-n] \Gamma[-3+\beta+n] / \Gamma[\beta]$ with $\beta = -2(9w_{\text{eff}} - 2p + 3)/(1 + 3w_{\text{eff}})$. The spectrum C_ℓ is shown in Fig. 16.2 for some values of n .

14.2 Let us expand a plane wave in Legendre polynomials

$$e^{ikr\mu_{kr}} = \sum_0^\infty (2\ell + 1) i^\ell \mathcal{P}_\ell(\mu_{kr}) j_\ell(kr), \quad (16.124)$$

where μ_{kr} denotes the cosine of the angle between $\mathbf{k}$ and $\mathbf{r}$. Equation (14.51) can be written as

$$\begin{aligned}\xi_s(\mathbf{r}) &= \int P_r(k)(1 + 2\beta\mu_{kd}^2 + \beta^2\mu_{kd}^4) e^{i\mathbf{k}\cdot\mathbf{r}} d^3k \\ &= \xi_r(\mathbf{r}) + \Xi_1 + \Xi_2,\end{aligned}\quad (16.125)$$

where μ_{kd} denotes the angle between $\mathbf{r}$ and the line of sight $\mathbf{d}$ and

$$\Xi_1 \equiv 2\beta \int P_r(k)\mu_{kd}^2 e^{i\mathbf{k}\cdot\mathbf{r}} d^3k, \quad \Xi_2 \equiv \beta^2 \int P_r(k)\mu_{kd}^4 e^{i\mathbf{k}\cdot\mathbf{r}} d^3k. \quad (16.126)$$

Now μ_{kd}^2 in Ξ_1 can be expanded in Legendre polynomials as

$$\mu_{kd}^2 = \sum_{\ell=0}^{\infty} b_{\ell} \mathcal{P}_{\ell} = \frac{1}{3} \mathcal{P}_0(\mu_{kd}) + \frac{2}{3} \mathcal{P}_2(\mu_{kd}), \quad (16.127)$$

and similarly μ_{kd}^4 in Ξ_2 :

$$\mu_{kd}^4 = \frac{1}{5} \mathcal{P}_0 + \frac{4}{7} \mathcal{P}_2 + \frac{8}{35} \mathcal{P}_4. \quad (16.128)$$

Expanding $e^{i\mathbf{k}\cdot\mathbf{r}}$, we then have

$$\Xi_1 = 2\beta \int P_r(k) \left[\frac{1}{3} \mathcal{P}_0(\mu_{kd}) + \frac{2}{3} \mathcal{P}_2(\mu_{kd}) \right] \left[\sum_{\ell=0}^{\infty} (2\ell+1) i^{\ell} \mathcal{P}_{\ell}(\mu_{kr}) j_{\ell}(kr) \right] k^2 dk d\Omega_k, \quad (16.129)$$

where $d\Omega_k$ denotes the integration over the angular coordinates of $\mathbf{k}$. Now we need to use the generalization of the orthonormality relation given by Eq. (14.144):

$$\int d\Omega_{kr} \mathcal{P}_{\ell'}(\mu_{kd}) \mathcal{P}_{\ell}(\mu_{kr}) = \frac{4\pi}{2\ell+1} \mathcal{P}_{\ell}(\mu_{rd}) \delta_{\ell\ell'}. \quad (16.130)$$

Then we obtain

$$\Xi_1 = 8\pi\beta \int P_r(k) \left[\frac{1}{3} \mathcal{P}_0(\mu_{rd}) j_0(kr) - \frac{2}{3} \mathcal{P}_2(\mu_{rd}) j_2(kr) \right] k^2 dk, \quad (16.131)$$

and similarly

$$\Xi_2 = 4\pi\beta^2 \int P_r(k) \left[\frac{1}{5} \mathcal{P}_0(\mu_{rd}) j_0(kr) - \frac{4}{7} \mathcal{P}_2(\mu_{rd}) j_2(kr) + \frac{8}{35} \mathcal{P}_4(\mu_{rd}) j_4(kr) \right] k^2 dk. \quad (16.132)$$

The integrals on $j_0 = \sin(x)/x$ give the isotropic correlation function

$$\xi_r(r) = 4\pi \int P_r(k) j_0(kr) k^2 dk. \quad (16.133)$$

For instance in Ξ_1 we have

$$8\pi\beta \int P_r(k) \frac{1}{3} \mathcal{P}_0(\mu_{rd}) j_0(kr) k^2 dk = \frac{8\pi}{3} \beta \int P_r(k) \frac{\sin(kr)}{kr} k^2 dk = \frac{2}{3} \beta \xi_r(r). \quad (16.134)$$

Let us consider now the term in Ξ_1 :

$$A \equiv -\frac{4}{3}\beta\mathcal{P}_2(\mu_{rd}) \cdot 4\pi \int P_r(k)j_2(kr)k^2dk. \quad (16.135)$$

We want to express $j_2(kr)$ in terms of $j_0(kr)$, so we can write I_2 in terms of the isotropic correlation function ξ_r . To do so we need to notice that

$$\begin{aligned} I_2(r) &\equiv \frac{3}{r^3} \int_0^r \xi_r(y)y^2dy = \frac{4\pi}{r^3} \int P_r(k)k^2dk \int_0^r y^2dy \frac{3 \sin(ky)}{ky} \\ &= 4\pi \int P_r(k) \frac{k^2}{(kr)^3} dk \int_0^{kr} \frac{3 \sin(z)}{z} z^2 dz \\ &= 4\pi \int P_r(k) \frac{k^2}{(kr)^3} dk (kr)^3 [j_2(kr) + j_0(kr)] \\ &= 4\pi \int P_r(k)k^2dk [j_2(kr) + j_0(kr)] \\ &= \xi_r(r) + 4\pi \int P_r(k)k^2dk j_2(kr), \end{aligned} \quad (16.136)$$

so that

$$A = -\frac{4}{3}\beta\mathcal{P}_2(\mu_{rd})[I_2(r) - \xi_r(r)]. \quad (16.137)$$

Analogously, we can use the relation

$$\begin{aligned} I_4 &\equiv \frac{5}{r^5} \int_0^r \xi_r(y)y^4dy = \frac{4\pi}{r^5} \int P_r(k)k^2dk \int_0^r y^4dy \frac{5 \sin(ky)}{ky} \\ &= 4\pi \int P_r(k) \frac{k^2}{(kr)^5} dk \int_0^{kr} \frac{5 \sin(z)}{z} z^4 dz \\ &= 4\pi \int P_r(k) \frac{k^2}{(kr)^5} dk (kr)^5 \left[-\frac{2}{7}j_4(kr) + \frac{5}{7}j_2(kr) + j_0(kr) \right] \\ &= \xi_r(r) + \frac{5}{7}[I_2 - \xi_r(r)] - \frac{8}{7}\pi \int P_r(k)k^2j_4(kr)dk, \end{aligned} \quad (16.138)$$

to simplify the j_4 term in Ξ_2 :

$$\begin{aligned} B &\equiv \frac{8}{35}\beta^2\mathcal{P}_4(\mu_{rd}) \cdot 4\pi \int P_r(k)j_4(kr)k^2dk \\ &= \frac{8}{35}\beta^2\mathcal{P}_4(\mu_{rd}) \left[\frac{7}{2}\xi_r + \frac{5}{2}(I_2 - \xi_r) - \frac{7}{2}I_4 \right]. \end{aligned} \quad (16.139)$$

Then we have

$$\Xi_1 = \frac{2}{3}\beta\xi_r - \frac{4}{3}\beta\mathcal{P}_2(\mu_{rd})[I_2(r) - \xi_r(r)], \quad (16.140)$$

$$\Xi_2 = \frac{1}{5}\beta^2\xi_r - \frac{4}{7}\beta^2\mathcal{P}_2(\mu_{rd})[I_2(r) - \xi_r(r)] + \frac{8}{35}\beta^2\mathcal{P}_4(\mu_{rd})\left(\xi_r + \frac{5}{2}I_2 - \frac{7}{2}I_4\right). \quad (16.141)$$

So finally the expansion Eq. (14.52) in multipoles $\ell = 0, 2, 4$ is

$$\begin{aligned} \xi_s = \mathcal{P}_0(\mu_{rd})\left(1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2\right)\xi_r + \mathcal{P}_2(\mu_{rd})\left(\frac{4}{3}\beta + \frac{4}{7}\beta^2\right)[\xi_r(r) - I_2(r)] \\ + \frac{8}{35}\beta^2\mathcal{P}_4(\mu_{rd})\left(\xi_r + \frac{5}{2}I_2 - \frac{7}{2}I_4\right). \end{aligned} \quad (16.142)$$

14.3 Integrating the right-hand side of Eq. (14.59) by parts we obtain (replacing r with s for later convenience)

$$\Xi(\sigma) = 2 \int_{\sigma}^{\infty} \frac{s\xi(s)}{(s^2 - \sigma^2)^{1/2}} ds = [\xi(s)(s^2 - \sigma^2)^{1/2}]_{\sigma}^{\infty} - 2 \int_{\sigma}^{\infty} \xi'(s)(s^2 - \sigma^2)^{1/2} ds. \quad (16.143)$$

The boundary term vanishes if $\xi(s)$ decreases to 0 faster than $1/s$ for $s \rightarrow \infty$ and if it is not singular at $s = \sigma$. Differentiating Eq. (16.143) with respect to σ , we find

$$\Xi'(\sigma) = 2 \int_{\sigma}^{\infty} \frac{\xi'(s)\sigma}{(s^2 - \sigma^2)^{1/2}} ds. \quad (16.144)$$

Inserting this into the r.h.s. of Eq. (14.60) we obtain

$$\begin{aligned} \text{The r.h.s. of Eq. (14.60)} &= -\frac{2}{\pi} \int_r^{\infty} d\sigma \int_{\sigma}^{\infty} ds \frac{\xi'(s)\sigma}{(s^2 - \sigma^2)^{1/2}(\sigma^2 - r^2)^{1/2}} \\ &= -\frac{2}{\pi} \int_r^{\infty} ds \xi'(s) \int_r^s \frac{\sigma d\sigma}{(s^2 - \sigma^2)^{1/2}(\sigma^2 - r^2)^{1/2}} \\ &= -\int_r^{\infty} ds \xi'(s) = \xi(r), \end{aligned} \quad (16.145)$$

where we have assumed $\xi(\infty) = 0$. The change of the integral limits in the third line is possible because the two-dimensional domain of integration $s \in (\sigma, \infty) \cap \sigma \in (r, \infty)$ is identical to $\sigma \in (r, s) \cap s \in (r, \infty)$.

14.4 We need to evaluate the integrals

$$q_{ij}^{(\text{circ})} = \int d\theta_x d\theta_y (f + 2\kappa_{\text{wl}}\theta^2 f_{,\theta^2})\theta_i\theta_j, \quad (16.146)$$

$$q_{ij}^{(\text{ell})} = -2 \int d\theta_x d\theta_y (\gamma_1\theta_x^2 - \gamma_1\theta_y^2 + 2\gamma_2\theta_x\theta_y)f_{,\theta^2}\theta_i\theta_j. \quad (16.147)$$

Let us define

$$E_{nm} \equiv \int d\theta_x d\theta_y f_{,\theta^2} \theta_x^n \theta_y^m . \quad (16.148)$$

Since f is a function that depends only on θ^2 , we have $E_{nm} = 0$ if either n or m is odd and $E_{nm} = E_{mn}$. Then we are left with

$$q_{ij}^{(\text{ell})} = -2 \begin{pmatrix} \gamma_1(E_{40} - E_{22}) & 2\gamma_2 E_{22} \\ 2\gamma_2 E_{22} & \gamma_1(E_{04} - E_{22}) \end{pmatrix} . \quad (16.149)$$

We have also

$$q_{ij}^{(\text{circ})} = \begin{pmatrix} U & 0 \\ 0 & U \end{pmatrix} + \mathcal{O}(\kappa_{\text{wl}}) , \quad (16.150)$$

where $U \equiv \int d\theta_x d\theta_y f \theta_x^2 = \int d\theta_x d\theta_y f \theta_y^2$. Changing to polar coordinates $\theta_x = \theta \cos \alpha$, $\theta_y = \theta \sin \alpha$, integrating over α , and putting $v = \theta^2$, the quantity U can be written as

$$U = \frac{1}{2} \int d\theta_x d\theta_y f \theta^2 = \pi \int d\theta f \theta^3 = \frac{\pi}{2} \int dv f(v) v . \quad (16.151)$$

Then we can evaluate the ellipticity

$$\varepsilon_1 = \frac{q_{xx} - q_{yy}}{q_{xx} + q_{yy}} = -2\gamma_1 \frac{E_{40} - 2E_{22} + E_{04}}{2U} . \quad (16.152)$$

The numerator can be integrated by changing again to polar coordinates and then by carrying out the partial integration:

$$\begin{aligned} E_{40} - 2E_{22} + E_{04} &= \int d\theta_x d\theta_y f_{,\theta^2} (\theta_x^2 - \theta_y^2)^2 \\ &= \int \theta^5 f_{,\theta^2} d\theta \int d\alpha (\cos^2 \alpha - \sin^2 \alpha)^2 \\ &= \frac{\pi}{2} \int v^2 \frac{df}{dv} dv = -\pi \int f v dv = -2U . \end{aligned} \quad (16.153)$$

So finally we obtain

$$\varepsilon_1 = 2\gamma_1 . \quad (16.154)$$

The $\mathcal{O}(\kappa_{\text{wl}})$ term in $q_{ij}^{(\text{circ})}$ is clearly a second-order correction. Similarly, we can write

$$E_{22} = \int d\theta_x d\theta_y f_{,\theta^2} \theta_x^2 \theta_y^2 = \frac{\pi}{4} \int \theta^5 d\theta f_{,\theta^2} = -\frac{\pi}{4} \int f v dv = -\frac{U}{2} , \quad (16.155)$$

and hence

$$\varepsilon_2 = \frac{2q_{xy}}{q_{xx} + q_{yy}} = -8\gamma_2 \frac{E_{22}}{2U} = 2\gamma_2 . \quad (16.156)$$

14.5 Suppose we have a field $f(x, y, r)$ projected along the r -direction in flat space with some unit-normalized weight $w(r)$,

$$F(\theta_x, \theta_y) = \int_0^\infty dr w(r) f(\mathbf{r}). \quad (16.157)$$

In what follows we always assume small angular deviations from the survey reference direction $\boldsymbol{\theta}$ so that spatial vectors are written as $\mathbf{r} = (\theta_x r, \theta_y r, r)$. The angular correlation function of F is

$$\begin{aligned} w_F(\Delta\boldsymbol{\theta}) &= \langle F(\boldsymbol{\theta}) F(\boldsymbol{\theta} + \Delta\boldsymbol{\theta}) \rangle \\ &= \int dr dr' w(r) w(r') \langle f([\theta_x r, \theta_y r, r] f[(\theta_x + \Delta\theta_x) r', (\theta_y + \Delta\theta_y) r', r']) \rangle \\ &\simeq \int dr w^2(r) \int dr' \xi(\theta'_x r' - \theta_x r, \theta'_y r' - \theta_y r, r'; r) \\ &\simeq \int dr w^2(r) \int dr' \xi(\Delta\theta_x r, \Delta\theta_y r, r'; r), \end{aligned} \quad (16.158)$$

where $\xi(\mathbf{x}; y)$ is the spatial correlation function at line-of-sight distance y for points separated by $\mathbf{x}$. In Eq. (16.158) we have assumed that the spatial angular correlation ξ depends only on the separation between two points, that it rapidly drops for large $r' - r$ (hence we can approximate $\theta'_x r' - \theta_x r \simeq r \Delta\theta$), and that $w(r)$ is slowly varying compared to the scale of fluctuations so that $w(r') \simeq w(r)$.

Now we define the angular transform

$$F(\mathbf{q}) = \int d^2\theta F(\boldsymbol{\theta}) e^{-i\mathbf{q} \cdot \boldsymbol{\theta}}, \quad (16.159)$$

for which the analog of Eq. (3.26) holds

$$\langle F(\mathbf{q}) F^*(\mathbf{q}') \rangle = (2\pi)^2 \delta_D(\mathbf{q} - \mathbf{q}') P_F(\mathbf{q}), \quad (16.160)$$

(we neglect the volume factors) where P_F is the power spectrum of F and the Fourier transform of $w(\boldsymbol{\theta})$. Denoting the power spectrum of $f(\mathbf{r})$ as $p_f(\mathbf{k})$, we have

$$\begin{aligned} P_F(\mathbf{q}) &= \int d^2\theta w_F(\boldsymbol{\theta}) e^{-i\mathbf{q} \cdot \boldsymbol{\theta}} \\ &= \int d^2\theta \int dr w^2(r) \int dr' \xi(\theta_x r, \theta_y r, r'; r) e^{-i\mathbf{q} \cdot \boldsymbol{\theta}} \\ &= \int d^2\theta \int dr w^2(r) \int dr' \int d^3k p_f(\mathbf{k}; r) e^{i(k_x r' + k_x \theta_x r + k_y \theta_y r) - i\mathbf{q} \cdot \boldsymbol{\theta}} \\ &= \int d^3k p_f(\mathbf{k}; r) \int dr w^2(r) \int d^2\theta dr' e^{-i(q_x - k_x r) \theta_x - i(q_y - k_y r) \theta_y} e^{i k_r r'}. \end{aligned} \quad (16.161)$$

Notice that in the integration variables we denoted $\Delta\theta$ as θ . The last integral gives a Dirac function $\delta_D(k_x r - q_x, k_y r - q_y, 0)$ so that using the formula

$$\begin{aligned} & \int dk_x dk_y dk_z \delta_D(k_x r - q_x, k_y r - q_y, 0) f(k_x, k_y, k_z) \\ &= \int d(\tilde{k}_x) d(\tilde{k}_y) dk_z r^{-2} \delta_D(\tilde{k}_x - q_x, \tilde{k}_y - q_y, 0) f = r^{-2} f\left(\frac{q_x}{r}, \frac{q_y}{r}, 0\right), \end{aligned} \quad (16.162)$$

we obtain

$$P_F(\mathbf{q}) = \int dr \frac{w^2}{r^2} p_f\left(\frac{\mathbf{q}}{r}, 0; r\right). \quad (16.163)$$

Assuming that both spectra are isotropic, i.e. they depend only on the modulus of their argument, we obtain Eq. (14.75). In a non-flat space the result would be

$$P_F(q) = \int dz \frac{w(z)^2}{\chi(z)^2} p_f\left[\frac{q}{\chi(z)}; z\right], \quad (16.164)$$

where χ is the comoving distance defined in Eq. (2.57) and the selection function $w(z)$ is normalized to unity in dz .

Mathematical Appendix

Here we collect some of the mathematical relations and definitions we have used throughout the book.

Our convention for the Fourier transformation is

$$f(\mathbf{x}) = \frac{V}{(2\pi)^3} \int f_{\mathbf{k}} e^{i\mathbf{k} \cdot \mathbf{x}} d^3k, \quad (17.1)$$

$$f_{\mathbf{k}} = \frac{1}{V} \int f(\mathbf{x}) e^{-i\mathbf{k} \cdot \mathbf{x}} d^3x. \quad (17.2)$$

Dirac's delta function $\delta_D(\mathbf{x})$ is defined as

$$\delta_D(\mathbf{x}) = (2\pi)^{-3} \int e^{i\mathbf{k} \cdot \mathbf{x}} d^3k. \quad (17.3)$$

The expansion in Legendre polynomials of a function $\Theta(\mu)$ is

$$\Theta(\mu) = \sum_{\ell} (2\ell + 1) i^{\ell} \Theta_{\ell} \mathcal{P}_{\ell}(\mu), \quad (17.4)$$

where

$$\Theta_{\ell} \equiv \frac{1}{(-i)^{\ell}} \int_{-1}^1 \frac{d\mu}{2} \mathcal{P}_{\ell}(\mu) \Theta(\mu), \quad (17.5)$$

and $\mathcal{P}_{\ell}$ is the Legendre polynomial of order ℓ . The first polynomials are $\mathcal{P}_0(\mu) = 1$, $\mathcal{P}_1(\mu) = \mu$, and $\mathcal{P}_2(\mu) = (3\mu^2 - 1)/2$. They obey the normalization rule

$$\int_{-1}^1 \mathcal{P}_m(\mu) \mathcal{P}_n(\mu) d\mu = \frac{2}{2n + 1} \delta_{nm}. \quad (17.6)$$

A plane wave can be expanded in Legendre polynomials as follows:

$$e^{i\mathbf{k}r\mu_{kr}} = \sum_{\ell=0}^{\infty} (2\ell + 1) i^{\ell} \mathcal{P}_{\ell}(\mu_{kr}) j_{\ell}(kr), \quad (17.7)$$

where μ_{kr} denotes the cosine of the angle between $\mathbf{k}$ and $\mathbf{r}$. The first spherical Bessel functions are

$$j_0(x) = \sin(x)/x, \quad (17.8)$$

$$j_2(x) = -3 \cos(x)/x^2 + (3 - x^2) \sin(x)/x^3, \quad (17.9)$$

$$j_4(x) = 5(2x^2 - 21) \cos(x)/x^4 + (105 - 45x^2 + x^4) \sin(x)/x^5. \quad (17.10)$$

A useful integral is

$$\int_0^\infty j_\ell(x) dx = \frac{\sqrt{\pi}}{2} \frac{\Gamma[(\ell + 1)/2]}{\Gamma[(\ell + 2)/2]}. \quad (17.11)$$

Because of the addition theorem

$$\mathcal{P}_\ell(\mu_{kr}) = \frac{4\pi}{2\ell + 1} \sum_n Y_{n\ell}^*(\hat{\mathbf{k}}) Y_{n\ell}(\hat{\mathbf{r}}), \quad (17.12)$$

we have also

$$e^{i\mathbf{k} \cdot \mathbf{r}} = 4\pi \sum_{\ell n} i^\ell j_\ell(kr) Y_{n\ell}^*(\hat{\mathbf{k}}) Y_{n\ell}(\hat{\mathbf{r}}). \quad (17.13)$$

Another useful relation is a theorem on Legendre polynomials which generalizes the orthonormality relation. If we have two vectors $\mathbf{k}$, $\mathbf{r}$ and the line of sight $\mathbf{d}$ and denote with μ_{kd} the cosine of the angle between $\mathbf{k}$ and $\mathbf{d}$, and analogously the cosines μ_{kr} , μ_{rd} , then we have that the integration over the angular coordinates of $\mathbf{k}$ is

$$\int d\Omega_k \mathcal{P}_{\ell'}(\mu_{kd}) \mathcal{P}_\ell(\mu_{kr}) = \frac{4\pi}{2\ell + 1} \mathcal{P}_\ell(\mu_{rd}) \delta_{\ell\ell'}. \quad (17.14)$$

This implies also

$$\int d\Omega_k Y_{\ell'm}^*(\hat{\mathbf{k}}) \mathcal{P}_\ell(\mu_{kr}) = \frac{4\pi}{2\ell + 1} Y_{\ell m}(\hat{\mathbf{r}}) \delta_{\ell\ell'}. \quad (17.15)$$

The spherical harmonics are orthonormal functions, satisfying

$$\int d\Omega Y_{\ell m}(\hat{\mathbf{r}}) Y_{\ell' m'}^*(\hat{\mathbf{r}}) = \delta_{\ell\ell'} \delta_{mm'}. \quad (17.16)$$

References

- [1] A. G. Riess *et al.*, Observational evidence from supernovae for an accelerating universe and a cosmological constant, *Astron. J.* **116** (1998), 1009.
- [2] S. Perlmutter *et al.*, Measurements of Ω and Λ from 42 high-redshift supernovae, *Astrophys. J.* **517** (1999), 565.
- [3] V. Sahni and A. A. Starobinsky, The case for a positive cosmological Λ -term, *Int. J. Mod. Phys. D* **9** (2000), 373.
- [4] S. M. Carroll, The cosmological constant, *Living Rev. Rel.* **4** (2001), 1.
- [5] P. J. E. Peebles and B. Ratra, The cosmological constant and dark energy, *Rev. Mod. Phys.* **75** (2003), 559.
- [6] T. Padmanabhan, Cosmological constant: The weight of the vacuum, *Phys. Rept.* **380** (2003), 235.
- [7] E. J. Copeland, M. Sami, and S. Tsujikawa, Dynamics of dark energy, *Int. J. Mod. Phys. D* **15** (2006), 1753.
- [8] A. A. Starobinsky, A new type of isotropic cosmological models without singularity, *Phys. Lett. B* **91** (1980), 99.
- [9] D. Kazanas, Dynamics of the universe and spontaneous symmetry breaking, *Astrophys. J.* **241** (1980), L59.
- [10] A. H. Guth, The inflationary universe: a possible solution to the horizon and flatness problems, *Phys. Rev. D* **23** (1981), 347.
- [11] K. Sato, First order phase transition of a vacuum and expansion of the universe, *Mon. Not. Roy. Astron. Soc.* **195** (1981), 467.
- [12] G. F. Smoot *et al.*, Structure in the COBE differential microwave radiometer first year maps, *Astrophys. J.* **396** (1992), L1.
- [13] D. N. Spergel *et al.* [WMAP Collaboration], First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Determination of cosmological parameters, *Astrophys. J. Suppl.* **148** (2003), 175.
- [14] D. N. Spergel *et al.* [WMAP Collaboration], Three-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Implications for cosmology, *Astrophys. J. Suppl.* **170** (2007), 377.
- [15] E. Komatsu *et al.* [WMAP Collaboration], Five-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Cosmological interpretation, *Astrophys. J. Suppl.* **180** (2009), 330.
- [16] C. Wetterich, An asymptotically vanishing time-dependent cosmological constant, *Astron. Astrophys.* **301** (1995), 321.
- [17] L. Amendola, Coupled quintessence, *Phys. Rev. D* **62** (2000), 043511.

- [18] G. Jungman, M. Kamionkowski, and K. Griest, Supersymmetric dark matter, *Phys. Rept.* **267** (1996), 195.
- [19] G. B. Gelmini, Search for dark matter, *Int. J. Mod. Phys. A* **23** (2008), 4273.
- [20] N. Fornengo, Status and perspectives of indirect and direct dark matter searches, *Adv. Space Res.* **41** (2008), 2010.
- [21] G. Bertone, Dark matter: A multidisciplinary approach, arXiv:astro-ph/0710.5603 (2007).
- [22] S. Weinberg, The cosmological constant problem, *Rev. Mod. Phys.* **61** (1989), 1.
- [23] A. Einstein, Cosmological considerations in the General Theory of Relativity, *Sitzungsber. Preuss. Akad. Wiss. Phys.-math. Klasse VI* (1917), 142.
- [24] S. Kachru, R. Kallosh, A. Linde, and S. P. Trivedi, De Sitter vacua in string theory, *Phys. Rev. D* **68** (2003), 046005.
- [25] L. Susskind, The anthropic landscape of string theory, arXiv:hep-th/0302219 (2003).
- [26] Y. Fujii, Origin of the gravitational constant and particle masses in scale invariant scalar-tensor theory, *Phys. Rev. D* **26** (1982), 2580.
- [27] R. D. Peccei, J. Sola, and C. Wetterich, Adjusting the cosmological constant dynamically: cosmons and a new force weaker than gravity, *Phys. Lett. B* **195** (1987), 183.
- [28] L. H. Ford, Cosmological constant damping by unstable scalar fields, *Phys. Rev. D* **35** (1987), 2339.
- [29] C. Wetterich, Cosmology and the fate of dilatation symmetry, *Nucl. Phys. B* **302** (1988), 668.
- [30] B. Ratra and P. J. E. Peebles, Cosmological consequences of a rolling homogeneous scalar field, *Phys. Rev. D* **37** (1988), 3406.
- [31] Y. Fujii and T. Nishioka, Model of a decaying cosmological constant, *Phys. Rev. D* **42** (1990), 361.
- [32] T. Chiba, N. Sugiyama, and T. Nakamura, Cosmology with x-matter, *Mon. Not. Roy. Astron. Soc.* **289** (1997), L5.
- [33] P. G. Ferreira and M. Joyce, Structure formation with a self-tuning scalar field, *Phys. Rev. Lett.* **79** (1997), 4740.
- [34] P. G. Ferreira and M. Joyce, Cosmology with a primordial scaling field, *Phys. Rev. D* **58** (1998), 023503.
- [35] R. R. Caldwell, R. Dave, and P. J. Steinhardt, Cosmological imprint of an energy component with general equation-of-state, *Phys. Rev. Lett.* **80** (1998), 1582.
- [36] S. M. Carroll, Quintessence and the rest of the world, *Phys. Rev. Lett.* **81** (1998), 3067.
- [37] E. J. Copeland, A. R. Liddle, and D. Wands, Exponential potentials and cosmological scaling solutions, *Phys. Rev. D* **57** (1998), 4686.
- [38] I. Zlatev, L. M. Wang, and P. J. Steinhardt, Quintessence, cosmic coincidence, and the cosmological constant, *Phys. Rev. Lett.* **82** (1999), 896.
- [39] P. J. Steinhardt, L. M. Wang, and I. Zlatev, Cosmological tracking solutions, *Phys. Rev. D* **59** (1999), 123504.
- [40] A. Hebecker and C. Wetterich, Quintessential adjustment of the cosmological constant, *Phys. Rev. Lett.* **85** (2000), 3339.
- [41] A. Hebecker and C. Wetterich, Natural quintessence?, *Phys. Lett. B* **497** (2001), 281.
- [42] T. Chiba, T. Okabe, and M. Yamaguchi, Kinetically driven quintessence, *Phys. Rev. D* **62** (2000), 023511.

- [43] C. Armendariz-Picon, V. F. Mukhanov, and P. J. Steinhardt, A dynamical solution to the problem of a small cosmological constant and late-time cosmic acceleration, *Phys. Rev. Lett.* **85** (2000), 4438.
- [44] C. Armendariz-Picon, V. F. Mukhanov, and P. J. Steinhardt, Essentials of k-essence, *Phys. Rev. D* **63** (2001), 103510.
- [45] A. Y. Kamenshchik, U. Moschella, and V. Pasquier, An alternative to quintessence, *Phys. Lett. B* **511** (2001), 265.
- [46] M. C. Bento, O. Bertolami, and A. A. Sen, Generalized Chaplygin gas, accelerated expansion and dark energy-matter unification, *Phys. Rev. D* **66** (2002), 043507.
- [47] J. A. Frieman, C. T. Hill, A. Stebbins, and I. Waga, Cosmology with ultralight pseudo Nambu-Goldstone bosons, *Phys. Rev. Lett.* **75** (1995), 2077.
- [48] P. Binetruy, Models of dynamical supersymmetry breaking and quintessence, *Phys. Rev. D* **60** (1999), 063502.
- [49] A. Masiero, M. Pietroni, and F. Rosati, SUSY QCD and quintessence, *Phys. Rev. D* **61** (2000), 023504.
- [50] P. Brax and J. Martin, Quintessence and supergravity, *Phys. Lett. B* **468** (1999), 40.
- [51] P. Brax and J. Martin, The robustness of quintessence, *Phys. Rev. D* **61** (2000), 103502.
- [52] E. J. Copeland, N. J. Nunes, and F. Rosati, Quintessence models in supergravity, *Phys. Rev. D* **62** (2000), 123503.
- [53] S. Capozziello, Curvature quintessence, *Int. J. Mod. Phys. D* **11** (2002), 483.
- [54] S. Capozziello, V. F. Cardone, S. Carloni, and A. Troisi, Curvature quintessence matched with observational data, *Int. J. Mod. Phys. D* **12** (2003), 1969.
- [55] S. M. Carroll, V. Duvvuri, M. Trodden, and M. S. Turner, Is cosmic speed-up due to new gravitational physics?, *Phys. Rev. D* **70** (2004), 043528.
- [56] L. Amendola, Scaling solutions in general non-minimal coupling theories, *Phys. Rev. D* **60** (1999), 043501.
- [57] J. P. Uzan, Cosmological scaling solutions of non-minimally coupled scalar fields, *Phys. Rev. D* **59** (1999), 123510.
- [58] T. Chiba, Quintessence, the gravitational constant, and gravity, *Phys. Rev. D* **60** (1999), 083508.
- [59] N. Bartolo and M. Pietroni, Scalar tensor gravity and quintessence, *Phys. Rev. D* **61** (2000), 023518.
- [60] F. Perrotta, C. Baccigalupi, and S. Matarrese, Extended quintessence, *Phys. Rev. D* **61** (2000), 023507.
- [61] G. R. Dvali, G. Gabadadze, and M. Porrati, 4D gravity on a brane in 5D Minkowski space, *Phys. Lett. B* **485** (2000), 208.
- [62] V. Sahni and Y. Shtanov, Braneworld models of dark energy, *JCAP* **0311** (2003), 014.
- [63] C. Brans and R. H. Dicke, Mach's principle and a relativistic theory of gravitation, *Phys. Rev.* **124** (1961), 925.
- [64] M. Gasperini and G. Veneziano, The pre-big bang scenario in string cosmology, *Phys. Rept.* **373** (2003), 1.
- [65] M. Tegmark *et al.* [SDSS Collaboration], Cosmological parameters from SDSS and WMAP, *Phys. Rev. D* **69** (2004), 103501.
- [66] U. Seljak *et al.* [SDSS Collaboration], Cosmological parameter analysis including SDSS Ly-alpha forest and galaxy bias: Constraints on the primordial spectrum of fluctuations, neutrino mass, and dark energy, *Phys. Rev. D* **71** (2005), 103515.
- [67] M. Tegmark *et al.* [SDSS Collaboration], Cosmological Constraints from the SDSS Luminous Red Galaxies, *Phys. Rev. D* **74** (2006), 123507.

- [68] D. J. Eisenstein *et al.* [SDSS Collaboration], Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies, *Astrophys. J.* **633** (2005), 560.
- [69] S. Weinberg, *Gravitation and Cosmology*, Wiley and Sons, New York (1972).
- [70] A. R. Liddle and D. H. Lyth, *Cosmological Inflation and Large-Scale Structure*, Cambridge University Press (2000).
- [71] E. Hubble, A relation between distance and radial velocity among extra-galactic nebulae, *Proc. Nat. Acad. Sci.* **15** (1929), 168.
- [72] W. L. Freedman *et al.* [HST Collaboration], Final results from the Hubble space telescope key project to measure the Hubble constant, *Astrophys. J.* **553** (2001), 47.
- [73] E. Kolb and M. Turner, *The Early Universe*, Addison-Wesley, Redwood City, CA (1990).
- [74] S. Dodelson, *Modern Cosmology*, Academic Press (Elsevier), New York (2003).
- [75] D. J. Fixsen, E. S. Cheng, J. M. Gales, J. C. Mather, R. A. Shafer, and E. L. Wright, The cosmic microwave background spectrum from the full COBE/FIRAS data set, *Astrophys. J.* **473** (1996), 576.
- [76] G. Mangano, G. Miele, S. Pastor, and M. Peloso, A precision calculation of the effective number of cosmological neutrinos, *Phys. Lett. B* **534** (2002), 8.
- [77] D. Tytler, J. M. O'Meara, N. Suzuki, and D. Lubin, Review of big bang nucleosynthesis and primordial abundances, *Phys. Scripta* **T85** (2000), 12.
- [78] S. Burles, K. M. Nollett, and M. S. Turner, Big-bang nucleosynthesis predictions for precision cosmology, *Astrophys. J.* **552** (2001), L1.
- [79] R. D. Peccei and H. R. Quinn, CP conservation in the presence of instantons, *Phys. Rev. Lett.* **38** (1977), 1440.
- [80] I. M. H. Etherington, On the definition of distance in general relativity, *Gen. Rel. Grav.* **39** (2007), 1055.
- [81] K. Ichikawa and T. Takahashi, Dark energy evolution and the curvature of the universe from recent observations, *Phys. Rev. D* **73** (2006), 083526.
- [82] C. Clarkson, M. Cortes, and B. A. Bassett, Dynamical dark energy or simply cosmic curvature?, *JCAP* **0708** (2007), 011.
- [83] R. R. Caldwell, A phantom menace?, *Phys. Lett. B* **545** (2002), 23.
- [84] R. R. Caldwell, M. Kamionkowski, and N. N. Weinberg, Phantom energy and cosmic doomsday, *Phys. Rev. Lett.* **91** (2003), 071301.
- [85] S. D. Landy and A. S. Szalay, Bias and variance of angular correlation functions, *Astrophys. J.* **412** (1993), 64.
- [86] H. A. Feldman, N. Kaiser, and J. A. Peacock, Power spectrum analysis of three-dimensional redshift surveys, *Astrophys. J.* **426** (1994), 23.
- [87] U. Seljak *et al.* [SDSS Collaboration], Cosmological parameter analysis including SDSS Ly-alpha forest and galaxy bias: Constraints on the primordial spectrum of fluctuations, neutrino mass, and dark energy, *Phys. Rev. D* **71** (2005), 103515.
- [88] S. Cole *et al.* [The 2dFGRS Collaboration], The 2dF Galaxy Redshift Survey: Power-spectrum analysis of the final dataset and cosmological implications, *Mon. Not. Roy. Astron. Soc.* **362** (2005), 505.
- [89] J. M. Bardeen, Gauge invariant cosmological perturbations, *Phys. Rev. D* **22** (1980), 1882.
- [90] H. Kodama and M. Sasaki, Cosmological perturbation theory, *Prog. Theor. Phys. Suppl.* **78** (1984), 1.
- [91] G. F. R. Ellis and M. Bruni, Covariant and gauge invariant approach to cosmological density fluctuations, *Phys. Rev. D* **40** (1989), 1804.

- [92] V. F. Mukhanov, H. A. Feldman, and R. H. Brandenberger, Theory of cosmological perturbations. Part 1. Classical perturbations. Part 2. Quantum theory of perturbations. Part 3. Extensions, *Phys. Rept.* **215** (1992), 203.
- [93] J. c. Hwang and H. r. Noh, Gauge-ready formulation of the cosmological kinetic theory in generalized gravity theories, *Phys. Rev. D* **65** (2002), 023512.
- [94] B. A. Bassett, S. Tsujikawa, and D. Wands, Inflation dynamics and reheating, *Rev. Mod. Phys.* **78** (2006), 537.
- [95] J. Peebles, *The Large-Scale Structure of the Universe*, Princeton University Press (1980).
- [96] M. Postman and T. R. Lauer, Brightest cluster galaxies as standard candles, *Astrophys. J.* **440** (1995), 28.
- [97] D. Sarkar, H. A. Feldman, and R. Watkins, Bulk flows from velocity field surveys: a consistency check, *Mon. Not. Roy. Astron. Soc.* **375** (2007), 691.
- [98] N. Kaiser, Clustering in real space and in redshift space, *Mon. Not. Roy. Astron. Soc.* **227** (1987), 1.
- [99] J. A. Peacock and S. J. Dodds, Non-linear evolution of cosmological power spectra, *Mon. Not. Roy. Astron. Soc.* **280** (1996), L19.
- [100] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, The statistics of peaks of Gaussian random fields, *Astrophys. J.* **304** (1986), 15.
- [101] N. Kaiser, On the spatial correlations of Abell clusters, *Astrophys. J.* **284** (1984), L9.
- [102] E. Carretta, R. G. Gratton, G. Clementini, and F. Fusi Pecci, Distances, ages and epoch of formation of globular clusters, *Astrophys. J.* **533** (2000), 215.
- [103] R. Jimenez, P. Thejll, U. Jorgensen, J. MacDonald, and B. Pagel, Ages of globular clusters: a new approach, *Mon. Not. Roy. Astron. Soc.* **282** (1996), 926.
- [104] B. M. S. Hansen *et al.*, The white dwarf cooling sequence of the globular cluster Messier 4, *Astrophys. J.* **574** (2002), L155.
- [105] S. Chandrasekhar, The maximum mass of ideal white dwarfs, *Astrophys. J.* **74** (1931), 81.
- [106] M. Hamuy *et al.*, The absolute luminosities of the Calan/Tololo Type Ia supernovae, *Astron. J.* **112** (1996), 2391.
- [107] P. Astier *et al.* [The SNLS Collaboration], The Supernova Legacy Survey: Measurement of Ω_M , Ω_Λ and w from the first year data set, *Astron. Astrophys.* **447** (2006), 31.
- [108] A. G. Riess *et al.* [Supernova Search Team Collaboration], Type Ia Supernova discoveries at $z > 1$ from the Hubble Space Telescope: Evidence for past deceleration and constraints on dark energy evolution, *Astrophys. J.* **607** (2004), 665.
- [109] A. 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, *Astrophys. J.* **659** (2007), 98.
- [110] W. M. Wood-Vasey *et al.* [ESSENCE Collaboration], Observational constraints on the nature of the dark energy: First cosmological results from the ESSENCE Supernova Survey, *Astrophys. J.* **666** (2007), 694.
- [111] T. M. Davis *et al.*, Scrutinizing exotic cosmological models using ESSENCE supernova data combined with other cosmological probes, *Astrophys. J.* **666** (2007), 716.
- [112] M. Kowalski *et al.*, Improved cosmological constraints from new, old and combined supernova datasets, *Astrophys. J.* **686** (2008), 749.

- [113] A. A. Penzias and R. W. Wilson, A measurement of excess antenna temperature at 4080-Mc/s, *Astrophys. J.* **142** (1965), 419.
- [114] P. de Bernardis *et al.* [Boomerang Collaboration], A flat universe from high-resolution maps of the cosmic microwave background radiation, *Nature* **404** (2000), 955.
- [115] A. Balbi *et al.*, Constraints on cosmological parameters from MAXIMA-1, *Astrophys. J.* **545** (2000) L1 [Erratum-ibid. **558** (2001), L145].
- [116] W. Hu and N. Sugiyama, Anisotropies in the cosmic microwave background: An analytic approach, *Astrophys. J.* **444** (1995), 489.
- [117] W. Hu and N. Sugiyama, Small scale cosmological perturbations: An analytic approach, *Astrophys. J.* **471** (1996), 542.
- [118] R. K. Sachs and A. M. Wolfe, Perturbations of a cosmological model and angular variations of the microwave background, *Astrophys. J.* **147** (1967), 73.
- [119] M. Doran and M. Lilley, The location of CMB peaks in a universe with dark energy, *Mon. Not. Roy. Astron. Soc.* **330** (2002), 965.
- [120] D. J. Eisenstein and W. Hu, Baryonic features in the matter transfer function, *Astrophys. J.* **496** (1998), 605.
- [121] A. Blanchard, M. Douspis, M. Rowan-Robinson, and S. Sarkar, Large-scale galaxy correlations as a test for dark energy, *Astron. Astrophys.* **449** (2006), 925.
- [122] M. Shoji, D. Jeong, and E. Komatsu, Extracting angular diameter distance and expansion rate of the universe from two-dimensional galaxy power spectrum at high redshifts: Baryon acoustic oscillation fitting versus full modeling, *Astrophys. J.* **693** (2009), 1404.
- [123] T. Okumura, T. Matsubara, D. J. Eisenstein, I. Kayo, C. Hikage, A. S. Szalay, and D. P. Schneider, Large-scale anisotropic correlation function of SDSS luminous red galaxies, *Astrophys. J.* **677** (2008), 889.
- [124] W. J. Percival *et al.*, Measuring the baryon acoustic oscillation scale using the SDSS and 2dFGRS, *Mon. Not. Roy. Astron. Soc.* **381** (2007), 1053.
- [125] T. Nishimichi *et al.*, Modeling non-linear evolution of baryon acoustic oscillations: Convergence regime of N-body simulations and analytic models, arXiv:astro-ph/0810.0813 (2008).
- [126] R. Takahashi *et al.*, Simulations of baryon acoustic oscillations II: Covariance matrix of the matter power spectrum, *Astrophys. J.* **700** (2009), 479.
- [127] M. Tegmark *et al.* [SDSS Collaboration], Cosmological constraints from the SDSS luminous red galaxies, *Phys. Rev. D* **74** (2006), 123507.
- [128] A. Einstein, The foundation of the general theory of relativity, *Annalen Phys.* **49** (1916), 769.
- [129] W. de Sitter, On the relativity of inertia: Remarks concerning Einstein's latest hypothesis, *Proc. Kon. Ned. Acad. Wet.* **19** (1917), 1217.
- [130] A. Friedmann, On the possibility of a world with constant negative curvature of space, *Z. Phys.* **21** (1924), 326.
- [131] G. Lemaître, A homogeneous universe of constant mass and growing radius accounting for the radial velocity of extragalactic nebulae, *Ann. Soc. Sci. Bruxelles*, Ser. 1, **47** (1927), 49.
- [132] A. Einstein, *The Meaning of Relativity*, Princeton University, Princeton (1945).
- [133] G. Gamov, *My World Line*, Viking, New York (1970).
- [134] B. Zumino, Supersymmetry and the vacuum, *Nucl. Phys.* **89** (1975), 535.
- [135] E. Bianchi and C. Rovelli, Why all these prejudices against a constant?, arXiv:1002.3966 [astro-ph.CO].
- [136] M. F. Sohnius, Introducing supersymmetry, *Phys. Rept.* **128** (1985), 39.

- [137] D. Bailin and A. Love, *Supersymmetric Gauge Field Theory and String Theory*, Institute of Physics Publishing (1994).
- [138] M. Green, J. H. Schwarz, and E. Witten, *Superstring Theory*, Cambridge University Press (1987).
- [139] M. T. Grisaru, W. Siegel, and M. Rocek, Improved methods for supergraphs, *Nucl. Phys. B* **159** (1979), 429.
- [140] E. Cremmer, S. Ferrara, C. Kounnas, and D. V. Nanopoulos, Naturally vanishing cosmological constant in $\mathcal{N} = 1$ Supergravity, *Phys. Lett. B* **133** (1983), 61.
- [141] M. Dine, R. Rohm, N. Seiberg, and E. Witten, Gluino condensation in superstring models, *Phys. Lett. B* **156** (1985), 55.
- [142] I. Affleck, M. Dine, and N. Seiberg, Dynamical supersymmetry breaking in supersymmetric QCD, *Nucl. Phys. B* **241** (1984), 493.
- [143] E. Witten, Dimensional reduction of superstring models, *Phys. Lett. B* **155** (1985), 151.
- [144] S. W. Hawking, The cosmological constant is probably zero, *Phys. Lett. B* **134** (1984), 403.
- [145] S. Kachru, M. B. Schulz, and E. Silverstein, Self-tuning flat domain walls in 5d gravity and string theory, *Phys. Rev. D* **62** (2000), 045021.
- [146] S. H. H. Tye and I. Wasserman, A brane world solution to the cosmological constant problem, *Phys. Rev. Lett.* **86** (2001), 1682.
- [147] J. Yokoyama, A cosmological constant from degenerate vacua, *Phys. Rev. Lett.* **88** (2002), 151302.
- [148] S. Mukohyama and L. Randall, A dynamical approach to the cosmological constant, *Phys. Rev. Lett.* **92** (2004), 211302.
- [149] G. L. Kane, M. J. Perry, and A. N. Zytlow, An approach to the cosmological constant problem(s), *Phys. Lett. B* **609** (2005), 7.
- [150] A. D. Dolgov and F. R. Urban, Dynamical vacuum energy via adjustment mechanism, *Phys. Rev. D* **77** (2008), 083503.
- [151] G. W. Gibbons, Aspects of supergravity theories, in *Supersymmetry, Supergravity, and Related Topics*, eds. F. del Aguila, J. A. de Azcarraga, and L. E. Ibanez, World Scientific (1985), 346–351.
- [152] J. Maldacena and C. Nunez, Supergravity description of field theories on curved manifolds and a no-go theorem, *Int. J. Mod. Phys. A* **16** (2001), 822.
- [153] J. Polchinski, Dirichlet-branes and Ramond-Ramond charges, *Phys. Rev. Lett.* **75** (1995), 4724.
- [154] P. Candelas and X. de la Ossa, Moduli space of Calabi-Yau manifolds, *Nucl. Phys. B* **355** (1991), 455.
- [155] S. Gukov, C. Vafa, and E. Witten, CFT's from Calabi-Yau four-folds, *Nucl. Phys. B* **584** (2000), 69.
- [156] E. Witten, Non-perturbative superpotentials in string theory, *Nucl. Phys. B* **474** (1996), 343.
- [157] S. B. Giddings, S. Kachru, and J. Polchinski, Hierarchies from fluxes in string compactifications, *Phys. Rev. D* **66** (2002), 106006.
- [158] S. Kachru, J. Pearson, and H. L. Verlinde, Brane/Flux annihilation and the string dual of a non-supersymmetric field theory, *JHEP* **0206** (2002), 021.
- [159] B. Carter, Large number coincidences and the anthropic principle in cosmology, IAU Symposium 63, in *Confrontation of Cosmological Theories with Observational Data*, Kluwer (1974), 291.
- [160] J. Barrow and F. Tipler, *The Cosmological Anthropic Principle*, Oxford University Press, Oxford (1988).

- [161] S. Weinberg, Anthropic bound on the cosmological constant, *Phys. Rev. Lett.* **59** (1987), 2607.
- [162] A. D. Linde, The new inflationary universe scenario, in *The Very Early Universe*, eds. G. W. Gibbons, S. W. Hawking, and S. Siklos, Cambridge University Press (1983), 205.
- [163] A. D. Linde, A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems, *Phys. Lett. B* **108** (1982), 389.
- [164] A. J. Albrecht and P. J. Steinhardt, Cosmology for Grand Unified Theories with radiatively induced symmetry breaking, *Phys. Rev. Lett.* **48** (1982), 1220.
- [165] A. D. Linde, The inflationary universe, *Rept. Prog. Phys.* **47** (1984), 925.
- [166] J. D. Brown and C. Teitelboim, Dynamical neutralization of the cosmological constant, *Phys. Lett. B* **195** (1987), 177.
- [167] J. D. Brown and C. Teitelboim, Neutralization of the cosmological constant by membrane creation, *Nucl. Phys. B* **297** (1988), 787.
- [168] L. F. Abbott, A mechanism for reducing the value of the cosmological constant, *Phys. Lett. B* **150** (1985), 427.
- [169] R. Bousso and J. Polchinski, Quantization of four-form fluxes and dynamical neutralization of the cosmological constant, *JHEP* **0006** (2000), 006.
- [170] F. Denef and M. R. Douglas, Distributions of flux vacua, *JHEP* **0405** (2004), 072.
- [171] R. Blumenhagen, F. Gmeiner, G. Honecker, D. Lust, and T. Weigand, The statistics of supersymmetric D-brane models, *Nucl. Phys. B* **713** (2005), 83.
- [172] J. Garriga and A. Vilenkin, A prescription for probabilities in eternal inflation, *Phys. Rev. D* **64** (2001), 023507.
- [173] R. Bousso, Holographic probabilities in eternal inflation, *Phys. Rev. Lett.* **97** (2006), 191302.
- [174] A. Linde, Towards a gauge invariant volume-weighted probability measure for eternal inflation, *JCAP* **0706** (2007), 017.
- [175] T. Padmanabhan, Dark energy: Mystery of the millennium, *AIP Conf. Proc.* **861** (2006), 179.
- [176] C. de Rham, G. Dvali, S. Hofmann, J. Khoury, O. Pujolas, M. Redi, and A. J. Tolley, Cascading gravity: Extending the Dvali-Gabadadze-Porrati model to higher dimension, *Phys. Rev. Lett.* **100** (2008), 251603.
- [177] C. de Rham, S. Hofmann, J. Khoury, and A. J. Tolley, Cascading gravity and degeneration, *JCAP* **0802** (2008), 011.
- [178] N. Afshordi, Gravitational Aether and the thermodynamic solution to the cosmological constant problem, arXiv:astro-ph/0807.2639 (2008).
- [179] T. Padmanabhan, Dark energy and gravity, *Gen. Rel. Grav.* **40** (2008), 529.
- [180] T. Padmanabhan, Dark energy and its implications for gravity, arXiv:gr-qc/0807.2356 (2008).
- [181] M. Kunz and D. Sapone, Dark energy versus modified gravity, *Phys. Rev. Lett.* **98** (2007), 121301.
- [182] R. R. Caldwell and E. V. Linder, The limits of quintessence, *Phys. Rev. Lett.* **95** (2005), 141301.
- [183] A. D. Linde, Chaotic inflation, *Phys. Lett. B* **129** (1983), 177.
- [184] A. D. Linde, Inflation and quantum cosmology, in *Three Hundred Years of Gravitation*, eds. S. W. Hawking and W. Israel, Cambridge University Press (1987), 604.
- [185] R. Kallosh, J. Kratochvil, A. Linde, E. V. Linder, and M. Shmakova, Observational bounds on cosmic doomsday, *JCAP* **0310** (2003), 015.

- [186] A. de la Macorra and G. Piccinelli, General scalar fields as quintessence, *Phys. Rev. D* **61** (2000), 123503.
- [187] S. C. C. Ng, N. J. Nunes, and F. Rosati, Applications of scalar attractor solutions to cosmology, *Phys. Rev. D* **64** (2001), 083510.
- [188] T. Barreiro, E. J. Copeland, and N. J. Nunes, Quintessence arising from exponential potentials, *Phys. Rev. D* **61** (2000), 127301.
- [189] V. Sahni and L. M. Wang, A new cosmological model of quintessence and dark matter, *Phys. Rev. D* **62** (2000), 103517.
- [190] A. R. Liddle, A. Mazumdar, and F. E. Schunck, Assisted inflation, *Phys. Rev. D* **58** (1998), 061301.
- [191] A. A. Coley and R. J. van den Hoogen, The dynamics of multi-scalar field cosmological models and assisted inflation, *Phys. Rev. D* **62** (2000), 023517.
- [192] S. A. Kim, A. R. Liddle, and S. Tsujikawa, Dynamics of assisted quintessence, *Phys. Rev. D* **72** (2005), 043506.
- [193] J. Ohashi and S. Tsujikawa, Assisted dark energy, *Phys. Rev. D* **80** (2009), 103513.
- [194] P. J. E. Peebles and A. Vilenkin, Quintessential inflation, *Phys. Rev. D* **59** (1999), 063505.
- [195] L. Kofman, A. D. Linde, and A. A. Starobinsky, Towards the theory of reheating after inflation, *Phys. Rev. D* **56** (1997), 3258.
- [196] G. N. Felder, L. Kofman, and A. D. Linde, Instant preheating, *Phys. Rev. D* **59** (1999), 123523.
- [197] R. Bean, S. H. Hansen, and A. Melchiorri, Early universe constraints on dark energy, *Phys. Rev. D* **64** (2001), 103508.
- [198] C. J. Copi, D. N. Schramm, and M. S. Turner, Assessing big bang nucleosynthesis, *Phys. Rev. Lett.* **75** (1995), 3981.
- [199] P. J. Kernan and S. Sarkar, No crisis for big bang nucleosynthesis, *Phys. Rev. D* **54** (1996), 3681.
- [200] C. F. Kolda and D. H. Lyth, Quintessential difficulties, *Phys. Lett. B* **458** (1999), 197.
- [201] T. R. Taylor, G. Veneziano, and S. Yankielowicz, Supersymmetric QCD and its massless limit: an effective Lagrangian analysis, *Nucl. Phys. B* **218** (1983), 493.
- [202] E. Witten, The cosmological constant from the viewpoint of string theory, arXiv:hep-ph/0002297 (2000).
- [203] R. Kallosh, A. D. Linde, S. Prokushkin, and M. Shmakova, Gauged supergravities, de Sitter space and cosmology, *Phys. Rev. D* **65** (2002), 105016.
- [204] R. Kallosh, A. Linde, S. Prokushkin, and M. Shmakova, Supergravity, dark energy and the fate of the universe, *Phys. Rev. D* **66** (2002), 123503.
- [205] P. Fre, M. Trigiante, and A. Van Proeyen, Stable de Sitter vacua from $\mathcal{N} = 2$ supergravity, *Class. Quant. Grav.* **19** (2002), 4167.
- [206] Y. Nomura, T. Watari, and T. Yanagida, Quintessence axion potential induced by electroweak instanton effects, *Phys. Lett. B* **484** (2000), 103.
- [207] K. Choi, String or M theory axion as a quintessence, *Phys. Rev. D* **62** (2000), 043509.
- [208] J. E. Kim and H. P. Nilles, A quintessential axion, *Phys. Lett. B* **553** (2003), 1.
- [209] L. J. Hall, Y. Nomura, and S. J. Oliver, Evolving dark energy with w not equal -1 , *Phys. Rev. Lett.* **95** (2005), 141302.
- [210] J. E. Lidsey, D. Wands, and E. J. Copeland, Superstring cosmology, *Phys. Rept.* **337** (2000), 343.
- [211] M. Gasperini, F. Piazza, and G. Veneziano, Quintessence as a run-away dilaton, *Phys. Rev. D* **65** (2002), 023508.

- [212] T. Damour, F. Piazza, and G. Veneziano, Runaway dilaton and equivalence principle violations, *Phys. Rev. Lett.* **89** (2002), 081601.
- [213] G. Veneziano, Large-N bounds on, and compositeness limit of, gauge and gravitational interactions, *JHEP* **0206** (2002), 051.
- [214] K. i. Maeda, Towards the Einstein-Hilbert action via conformal transformation, *Phys. Rev. D* **39** (1989), 3159.
- [215] A. A. Starobinsky, How to determine an effective potential for a variable cosmological term, *JETP Lett.* **68** (1998), 757.
- [216] D. Huterer and M. S. Turner, Prospects for probing the dark energy via supernova distance measurements, *Phys. Rev. D* **60** (1999), 081301.
- [217] T. Nakamura and T. Chiba, Determining the equation of state of the expanding universe: Inverse problem in cosmology, *Mon. Not. Roy. Astron. Soc.* **306** (1999), 696.
- [218] T. Chiba and T. Nakamura, Feasibility of reconstructing the quintessential potential using SN Ia data, *Phys. Rev. D* **62** (2000), 121301.
- [219] T. D. Saini, S. Raychaudhury, V. Sahni, and A. A. Starobinsky, Reconstructing the cosmic equation of state from Supernova distances, *Phys. Rev. Lett.* **85** (2000), 1162.
- [220] U. Alam, V. Sahni, T. D. Saini, and A. A. Starobinsky, Exploring the expanding universe and dark energy using the statefinder diagnostic, *Mon. Not. Roy. Astron. Soc.* **344** (2003), 1057.
- [221] U. Alam, V. Sahni, T. D. Saini, and A. A. Starobinsky, Is there Supernova evidence for dark energy metamorphosis?, *Mon. Not. Roy. Astron. Soc.* **354** (2004), 275.
- [222] G. Efstathiou, Constraining the equation of state of the universe from distant Type Ia Supernovae and Cosmic Microwave Background Anisotropies, *Mon. Not. Roy. Astron. Soc.* **342** (2000), 810.
- [223] M. Chevalier and D. Polarski, Accelerating universes with scaling dark matter, *Int. J. Mod. Phys. D* **10** (2001), 213.
- [224] J. Weller and A. J. Albrecht, Future Supernovae observations as a probe of dark energy, *Phys. Rev. D* **65** (2002), 103512.
- [225] I. Maor, R. Brustein, J. McMahon, and P. J. Steinhardt, Measuring the equation-of-state of the universe: Pitfalls and prospects, *Phys. Rev. D* **65** (2002), 123003.
- [226] B. A. Bassett, M. Kunz, J. Silk, and C. Ungarelli, A late-time transition in the cosmic dark energy?, *Mon. Not. Roy. Astron. Soc.* **336** (2002), 1217.
- [227] P. S. Corasaniti and E. J. Copeland, A model independent approach to the dark energy equation of state, *Phys. Rev. D* **67** (2003), 063521.
- [228] E. V. Linder, Exploring the expansion history of the universe, *Phys. Rev. Lett.* **90** (2003), 091301.
- [229] S. Nesseris and L. Perivolaropoulos, A comparison of cosmological models using recent supernova data, *Phys. Rev. D* **70** (2004), 043531.
- [230] R. Lazkoz, S. Nesseris, and L. Perivolaropoulos, Exploring cosmological expansion parametrizations with the gold SnIa dataset, *JCAP* **0511** (2005), 010.
- [231] H. K. Jassal, J. S. Bagla, and T. Padmanabhan, WMAP constraints on low redshift evolution of dark energy, *Mon. Not. Roy. Astron. Soc.* **356** (2005), L11.
- [232] A. Upadhye, M. Ishak, and P. J. Steinhardt, Dynamical dark energy: Current constraints and forecasts, *Phys. Rev. D* **72** (2005), 063501.
- [233] R. Crittenden, E. Majerotto, and F. Piazza, Measuring deviations from a cosmological constant: a field-space parameterization, *Phys. Rev. Lett.* **98** (2007), 251301.

- [234] K. Ichikawa and T. Takahashi, Dark energy parametrizations and the curvature of the universe, *JCAP* **0702** (2007), 001.
- [235] G. B. Zhao, J. Q. Xia, B. Feng, and X. Zhang, Probing dynamics of dark energy with supernova, galaxy clustering and the three-year Wilkinson Microwave Anisotropy Probe (WMAP) observations, *Int. J. Mod. Phys. D* **16** (2007), 1229.
- [236] V. Sahni, T. D. Saini, A. A. Starobinsky, and U. Alam, Statefinder – a new geometrical diagnostic of dark energy, *JETP Lett.* **77** (2003), 201.
- [237] V. Sahni and A. Starobinsky, Reconstructing dark energy, *Int. J. Mod. Phys. D* **15** (2006), 2105.
- [238] E. V. Linder, The paths of quintessence, *Phys. Rev. D* **73** (2006), 063010.
- [239] C. Armendariz-Picon, T. Damour, and V. F. Mukhanov, k-Inflation, *Phys. Lett. B* **458** (1999), 209.
- [240] N. Arkani-Hamed, H. C. Cheng, M. A. Luty, and S. Mukohyama, Ghost condensation and a consistent infrared modification of gravity, *JHEP* **0405** (2004), 074.
- [241] F. Piazza and S. Tsujikawa, Dilatonic ghost condensate as dark energy, *JCAP* **0407** (2004), 004.
- [242] M. R. Garousi, Tachyon couplings on non-BPS D-branes and Dirac-Born-Infeld action, *Nucl. Phys. B* **584** (2000), 284.
- [243] A. Sen, Rolling tachyon, *JHEP* **0204** (2002), 048.
- [244] D. Kutasov and V. Niarchos, Tachyon effective actions in open string theory, *Nucl. Phys. B* **666** (2003), 56.
- [245] T. Padmanabhan, Accelerated expansion of the universe driven by tachyonic matter, *Phys. Rev. D* **66** (2002), 021301.
- [246] L. R. W. Abramo and F. Finelli, Cosmological dynamics of the tachyon with an inverse power-law potential, *Phys. Lett. B* **575** (2003), 165.
- [247] J. M. Aguirregabiria and R. Lazkoz, Tracking solutions in tachyon cosmology, *Phys. Rev. D* **69** (2004), 123502.
- [248] E. J. Copeland, M. R. Garousi, M. Sami, and S. Tsujikawa, What is needed of a tachyon if it is to be the dark energy?, *Phys. Rev. D* **71** (2005), 043003.
- [249] M. R. Garousi, M. Sami, and S. Tsujikawa, Inflation and dark energy arising from rolling massive scalar field on the D-brane, *Phys. Rev. D* **70** (2004), 043536.
- [250] E. Silverstein and D. Tong, Scalar speed limits and cosmology: Acceleration from D-cceleration, *Phys. Rev. D* **70** (2004), 103505.
- [251] M. Alishahiha, E. Silverstein, and D. Tong, DBI in the sky, *Phys. Rev. D* **70** (2004), 123505.
- [252] J. M. Maldacena, The large N limit of superconformal field theories and supergravity, *Adv. Theor. Math. Phys.* **2** (1998), 231.
- [253] J. Martin and M. Yamaguchi, DBI-essence, *Phys. Rev. D* **77** (2008), 123508.
- [254] Z. K. Guo and N. Ohta, Cosmological evolution of Dirac-Born-Infeld field, *JCAP* **0804** (2008), 035.
- [255] J. Garriga and V. F. Mukhanov, Perturbations in k-inflation, *Phys. Lett. B* **458** (1999), 219.
- [256] S. M. Carroll, M. Hoffman, and M. Trodden, Can the dark energy equation-of-state parameter w be less than -1 ?, *Phys. Rev. D* **68** (2003), 023509.
- [257] J. M. Cline, S. Jeon, and G. D. Moore, The phantom menaced: Constraints on low-energy effective ghosts, *Phys. Rev. D* **70** (2004), 043543.
- [258] E. Babichev, V. Mukhanov, and A. Vikman, k-essence, superluminal propagation, causality and emergent geometry, *JHEP* **0802** (2008), 101.

- [259] M. Malquarti, E. J. Copeland, and A. R. Liddle, k-essence and the coincidence problem, *Phys. Rev. D* **68** (2003), 023512.
- [260] C. Bonvin, C. Caprini, and R. Durrer, A no-go theorem for k-essence dark energy, *Phys. Rev. Lett.* **97** (2006), 081303.
- [261] P. Singh, M. Sami, and N. Dadhich, Cosmological dynamics of phantom field, *Phys. Rev. D* **68** (2003), 023522.
- [262] M. Sami and A. Toporensky, Phantom field and the fate of universe, *Mod. Phys. Lett. A* **19** (2004), 1509.
- [263] P. Sreekumar *et al.* [EGRET Collaboration], EGRET observations of the extragalactic gamma ray emission, *Astrophys. J.* **494** (1998), 523.
- [264] R. Gannouji, D. Polarski, A. Ranquet, and A. A. Starobinsky, Scalar-tensor models of normal and phantom dark energy, *JCAP* **0609** (2006), 016.
- [265] B. Feng, X. L. Wang, and X. M. Zhang, Dark energy constraints from the cosmic age and supernova, *Phys. Lett. B* **607** (2005), 35.
- [266] Z. K. Guo, Y. S. Piao, X. M. Zhang, and Y. Z. Zhang, Cosmological evolution of a quintom model of dark energy, *Phys. Lett. B* **608** (2005), 177.
- [267] L. Amendola and S. Tsujikawa, Phantom crossing, equation-of-state singularities, and local gravity constraints in $f(R)$ models, *Phys. Lett. B* **660** (2008), 125.
- [268] L. Perivolaropoulos, Reconstruction of extended quintessence potentials from the SN Ia gold dataset, *JCAP* **0510** (2005), 001.
- [269] J. Martin, C. Schmid, and J. P. Uzan, Testing for $w < -1$ in the solar system, *Phys. Rev. Lett.* **96** (2006), 061303.
- [270] M. Libanov, V. Rubakov, E. Papantonopoulos, M. Sami, and S. Tsujikawa, UV stable, Lorentz-violating dark energy with transient phantom era, *JCAP* **0708** (2007), 010.
- [271] A. Fuzfa and J. M. Alimi, Toward a unified description of dark energy and dark matter from the abnormally weighting energy hypothesis, *Phys. Rev. D* **75** (2007), 123007.
- [272] L. Amendola, D. Polarski, and S. Tsujikawa, Are $f(R)$ dark energy models cosmologically viable?, *Phys. Rev. Lett.* **98** (2007), 131302.
- [273] S. Tsujikawa, K. Uddin, S. Mizuno, R. Tavakol, and J. Yokoyama, Constraints on scalar-tensor models of dark energy from observational and local gravity tests, *Phys. Rev. D* **77** (2008), 103009.
- [274] N. Dalal, K. Abazajian, E. E. Jenkins, and A. V. Manohar, Testing the cosmic coincidence problem and the nature of dark energy, *Phys. Rev. Lett.* **87** (2001), 141302.
- [275] W. Zimdahl, D. Pavon, and L. P. Chimento, Interacting quintessence, *Phys. Lett. B* **521** (2001), 133.
- [276] S. del Campo, R. Herrera, G. Olivares, and D. Pavon, Interacting models of soft coincidence, *Phys. Rev. D* **74** (2006), 023501.
- [277] H. Wei and S. N. Zhang, Observational $H(z)$ data and cosmological models, *Phys. Lett. B* **644** (2007), 7.
- [278] L. Amendola, G. C. Campos, and R. Rosenfeld, Consequences of dark matter-dark energy interaction on cosmological parameters derived from SN Ia data, *Phys. Rev. D* **75** (2007), 083506.
- [279] Z. K. Guo, N. Ohta, and S. Tsujikawa, Probing the coupling between dark components of the universe, *Phys. Rev. D* **76** (2007), 023508.
- [280] G. Caldera-Cabral, R. Maartens, and L. A. Urena-Lopez, Dynamics of interacting dark energy, *Phys. Rev. D* **79** (2009), 063518.

- [281] T. Damour, G. W. Gibbons, and C. Gundlach, Dark matter, time varying G , and a dilaton field, *Phys. Rev. Lett.* **64** (1990), 123.
- [282] A. Fuzfa and J. M. Alimi, Dark energy as a Born-Infeld gauge interaction violating the equivalence principle, *Phys. Rev. Lett.* **97** (2006), 061301.
- [283] B. Gumjudpai, T. Naskar, M. Sami, and S. Tsujikawa, Coupled dark energy: Towards a general description of the dynamics, *JCAP* **0506** (2005), 007.
- [284] L. Amendola and D. Tocchini-Valentini, Stationary dark energy: the present universe as a global attractor, *Phys. Rev. D* **64** (2001), 043509.
- [285] L. Amendola and C. Quercellini, Tracking and coupled dark energy as seen by WMAP, *Phys. Rev. D* **68** (2003), 023514.
- [286] G. La Vacca, J. R. Kristiansen, L. P. L. Colombo, R. Mainini, and S. A. Bonometto, Do WMAP data favor neutrino mass and a coupling between cold dark matter and dark energy?, *JCAP* **0904** (2009), 007.
- [287] J. P. Uzan, The fundamental constants and their variation: Observational status and theoretical motivations, *Rev. Mod. Phys.* **75** (2003), 403.
- [288] D. Tocchini-Valentini and L. Amendola, Stationary dark energy with a baryon dominated era: Solving the coincidence problem with a linear coupling, *Phys. Rev. D* **65** (2002), 063508.
- [289] L. Amendola, M. Gasperini, and F. Piazza, SNLS data are consistent with acceleration at $z = 3$, *Phys. Rev. D* **74** (2006), 127302.
- [290] J. Valiviita, E. Majerotto, and R. Maartens, Instability in interacting dark energy and dark matter fluids, *JCAP* **0807** (2008), 020.
- [291] P. Q. Hung, Sterile neutrino and accelerating universe, arXiv:hep-ph/0010126 (2000).
- [292] P. Gu, X. Wang, and X. Zhang, Dark energy and neutrino mass limits from baryogenesis, *Phys. Rev. D* **68** (2003), 087301.
- [293] R. Fardon, A. E. Nelson, and N. Weiner, Dark energy from mass varying neutrinos, *JCAP* **0410** (2004), 005.
- [294] R. Takahashi and M. Tanimoto, Model of mass varying neutrinos in SUSY, *Phys. Lett. B* **633** (2006), 675.
- [295] K. Bamba, C. Q. Geng, and S. H. Ho, Radiative neutrino mass generation and dark energy, *JCAP* **0809** (2008), 001.
- [296] R. D. Peccei, Neutrino models of dark energy, *Phys. Rev. D* **71** (2005), 023527.
- [297] N. Afshordi, M. Zaldarriaga, and K. Kohri, On the stability of dark energy with mass-varying neutrinos, *Phys. Rev. D* **72** (2005), 065024.
- [298] A. W. Brookfield, C. van de Bruck, D. F. Mota, and D. Tocchini-Valentini, Cosmology with massive neutrinos coupled to dark energy, *Phys. Rev. Lett.* **96** (2006), 061301.
- [299] A. W. Brookfield, C. van de Bruck, D. F. Mota, and D. Tocchini-Valentini, Cosmology of mass-varying neutrinos driven by quintessence: Theory and observations, *Phys. Rev. D* **73** (2006), 083515.
- [300] O. E. Bjaelde, A. W. Brookfield, C. van de Bruck, S. Hannestad, D. F. Mota, L. Schrempp, and D. Tocchini-Valentini, Neutrino dark energy – revisiting the stability issue, *JCAP* **0801** (2008), 026.
- [301] K. Ichiki and Y. Y. Keum, Neutrino masses from cosmological probes in interacting neutrino dark-energy models, *JHEP* **0806** (2008), 058.
- [302] L. Amendola, M. Baldi, and C. Wetterich, Growing matter, *Phys. Rev. D* **78** (2008), 023015.
- [303] D. F. Mota, V. Pettorino, G. Robbers, and C. Wetterich, Neutrino clustering in growing neutrino quintessence, *Phys. Lett. B* **663** (2008), 160.

- [304] Y. Fujii *et al.*, The nuclear interaction at Oklo 2 billion years ago, *Nucl. Phys. B* **573** (2000), 377.
- [305] M. T. Murphy *et al.*, Possible evidence for a variable fine structure constant from QSO absorption lines: motivations, analysis and results, *Mon. Not. Roy. Astron. Soc.* **327** (2001), 1208.
- [306] J. K. Webb *et al.*, Further evidence for cosmological evolution of the fine structure constant, *Phys. Rev. Lett.* **87** (2001), 091301.
- [307] H. Chand, R. Srianand, P. Petitjean, and B. Aracil, Probing the cosmological variation of the fine-structure constant: Results based on VLT-UVES sample, *Astron. Astrophys.* **417** (2004), 853.
- [308] M. T. Murphy, J. K. Webb, V. V. Flambaum, and S. J. Curran, Does the fine structure constant vary? A detailed investigation into systematic effects, *Astrophys. Space Sci.* **283** (2003), 577.
- [309] J. D. Bekenstein, Fine structure constant: Is it really a constant?, *Phys. Rev. D* **25** (1982), 1527.
- [310] H. B. Sandvik, J. D. Barrow, and J. Magueijo, A simple varying-alpha cosmology, *Phys. Rev. Lett.* **88** (2002), 031302.
- [311] G. R. Dvali and M. Zaldarriaga, Changing alpha with time: Implications for fifth-force-type experiments and quintessence, *Phys. Rev. Lett.* **88** (2002), 091303.
- [312] T. Chiba and K. Kohri, Quintessence cosmology and varying alpha, *Prog. Theor. Phys.* **107** (2002), 631.
- [313] D. Parkinson, B. A. Bassett, and J. D. Barrow, Mapping the dark energy with varying alpha, *Phys. Lett. B* **578** (2004), 235.
- [314] S. Lee, K. A. Olive, and M. Pospelov, Quintessence models and the cosmological evolution of alpha, *Phys. Rev. D* **70** (2004), 083503.
- [315] E. J. Copeland, N. J. Nunes, and M. Pospelov, Models of quintessence coupled to the electromagnetic field and the cosmological evolution of alpha, *Phys. Rev. D* **69** (2004), 023501.
- [316] D. F. Mota and J. D. Barrow, Varying alpha in a more realistic universe, *Phys. Lett. B* **581** (2004), 141.
- [317] H. Marion *et al.*, A search for variations of fundamental constants using atomic fountain clocks, *Phys. Rev. Lett.* **90** (2003), 150801.
- [318] T. Rosenband *et al.*, Frequency ratio of Al^+ and Hg^+ single-ion optical clocks; Metrology at the 17th decimal place, *Science* **319** (2008), 1808.
- [319] N. J. Nunes, T. Dent, C. J. A. P. Martins, and G. Robbers, Reconstructing the evolution of dark energy with variations of fundamental parameters, *Mem. SAI* **80** (2009), 785.
- [320] M. R. Garousi, M. Sami, and S. Tsujikawa, Constraints on Dirac-Born-Infeld type dark energy models from varying alpha, *Phys. Rev. D* **71** (2005), 083005.
- [321] J. Khoury and A. Weltman, Chameleon fields: Awaiting surprises for tests of gravity in space, *Phys. Rev. Lett.* **93** (2004), 171104.
- [322] J. Khoury and A. Weltman, Chameleon cosmology, *Phys. Rev. D* **69** (2004), 044026.
- [323] T. Tamaki and S. Tsujikawa, Revisiting chameleon gravity: Thin-shell and no-shell fields with appropriate boundary conditions, *Phys. Rev. D* **78** (2008), 084028.
- [324] S. Tsujikawa, T. Tamaki, and R. Tavakol, Chameleon scalar fields in relativistic gravitational backgrounds, *JCAP* **0905** (2009), 020.
- [325] C. M. Will, The confrontation between general relativity and experiment, *Living Rev. Rel.* **9** (2005), 3.

- [326] C. D. Hoyle *et al.*, Sub-millimeter tests of the gravitational inverse-square law, *Phys. Rev. D* **70** (2004), 042004.
- [327] P. Brax, C. van de Bruck, A. C. Davis, J. Khoury, and A. Weltman, Detecting dark energy in orbit: The cosmological chameleon, *Phys. Rev. D* **70** (2004), 123518.
- [328] D. F. Mota and D. J. Shaw, Strongly coupled chameleon fields: New horizons in scalar field theory, *Phys. Rev. Lett.* **97** (2006), 151102.
- [329] P. Brax, C. van de Bruck, A. C. Davis, D. F. Mota, and D. Shaw, Detecting chameleons through Casimir force measurements, *Phys. Rev. D* **76** (2007), 085010.
- [330] L. Amendola, M. Quartin, S. Tsujikawa, and I. Waga, Challenges for scaling cosmologies, *Phys. Rev. D* **74** (2006), 023525.
- [331] S. Tsujikawa and M. Sami, A unified approach to scaling solutions in a general cosmological background, *Phys. Lett. B* **603** (2004), 113.
- [332] S. Mizuno and K. i. Maeda, Quintessence in a brane world, *Phys. Rev. D* **64** (2001), 123521.
- [333] E. J. Copeland, S. J. Lee, J. E. Lidsey, and S. Mizuno, Generalized cosmological scaling solutions, *Phys. Rev. D* **71** (2005), 023526.
- [334] S. Tsujikawa, General analytic formulae for attractor solutions of scalar-field dark energy models and their multi-field generalizations, *Phys. Rev. D* **73** (2006), 103504.
- [335] R. J. Scherrer, Purely kinetic k-essence as unified dark matter, *Phys. Rev. Lett.* **93** (2004), 011301.
- [336] D. Carturan and F. Finelli, Cosmological effects of a class of fluid dark energy models, *Phys. Rev. D* **68** (2003), 103501.
- [337] H. Sandvik, M. Tegmark, M. Zaldarriaga, and I. Waga, The end of unified dark matter?, *Phys. Rev. D* **69** (2004), 123524.
- [338] P. P. Avelino *et al.*, The onset of the non-linear regime in unified dark matter models, *Phys. Rev. D* **69** (2004), 041301.
- [339] L. Amendola, F. Finelli, C. Burigana, and D. Carturan, WMAP and the generalized Chaplygin gas, *JCAP* **0307** (2003), 005.
- [340] L. Amendola, I. Waga, and F. Finelli, Observational constraints on silent quartessence, *JCAP* **0511** (2005), 009.
- [341] D. Pietrobon, A. Balbi, M. Bruni, and C. Quercellini, Affine parameterization of the dark sector: constraints from WMAP5 and SDSS, *Phys. Rev. D* **78** (2008), 083510.
- [342] D. Bertacca, S. Matarrese, and M. Pietroni, Unified dark matter in scalar field cosmologies, *Mod. Phys. Lett. A* **22** (2007), 2893.
- [343] T. Fukuyama, M. Morikawa, and T. Tatekawa, Cosmic structures via Bose Einstein condensation and its collapse, *JCAP* **0806** (2008), 033.
- [344] S. Nojiri, S. D. Odintsov, and S. Tsujikawa, Properties of singularities in (phantom) dark energy universe, *Phys. Rev. D* **71** (2005), 063004.
- [345] J. D. Barrow, Sudden future singularities, *Class. Quant. Grav.* **21** (2004), L79.
- [346] H. Stefancic, ‘Expansion’ around the vacuum equation of state: Sudden future singularities and asymptotic behavior, *Phys. Rev. D* **71** (2005), 084024.
- [347] I. H. Brevik and O. Gorbunova, Dark energy and viscous cosmology, *Gen. Rel. Grav.* **37** (2005), 2039.
- [348] M. P. Dabrowski, Statefinders, higher-order energy conditions and sudden future singularities, *Phys. Lett. B* **625** (2005), 184.
- [349] M. Bouhmadi-Lopez, P. F. Gonzalez-Diaz, and P. Martin-Moruno, Worse than a big rip?, *Phys. Lett. B* **659** (2008), 1.
- [350] J. D. Barrow, G. Galloway, and F. Tipler, The closed-universe recollapse conjecture, *Mon. Not. Roy. Astron. Soc.* **223** (1986), 835.

- [351] J. D. Barrow, Graduated inflationary universes, *Phys. Lett. B* **235** (1990), 40.
- [352] M. Sami, P. Singh, and S. Tsujikawa, Avoidance of future singularities in loop quantum cosmology, *Phys. Rev. D* **74** (2006), 043514.
- [353] D. Samart and B. Gumjudpai, Phantom field dynamics in loop quantum cosmology, *Phys. Rev. D* **76** (2007), 043514.
- [354] M. Bojowald, Loop quantum cosmology, *Living Rev. Rel.* **8** (2005), 11.
- [355] A. Ashtekar, T. Pawłowski, and P. Singh, Quantum nature of the big bang: An analytical and numerical investigation, *Phys. Rev. D* **73** (2006), 124038.
- [356] S. Nojiri and S. D. Odintsov, Modified gravity with negative and positive powers of the curvature: Unification of the inflation and of the cosmic acceleration, *Phys. Rev. D* **68** (2003), 123512.
- [357] T. Chiba, $1/R$ gravity and scalar-tensor gravity, *Phys. Lett. B* **575** (2003), 1.
- [358] A. D. Dolgov and M. Kawasaki, Can modified gravity explain accelerated cosmic expansion?, *Phys. Lett. B* **573**, 1 (2003).
- [359] M. E. Soussa and R. P. Woodard, The force of gravity from a Lagrangian containing inverse powers of the Ricci scalar, *Gen. Rel. Grav.* **36**, (2004), 855.
- [360] G. J. Olmo, Post-Newtonian constraints on $f(R)$ cosmologies in metric and Palatini formalism, *Phys. Rev. D* **72** (2005), 083505.
- [361] V. Faraoni, Solar system experiments do not yet veto modified gravity models, *Phys. Rev. D* **74** (2006), 023529.
- [362] I. Navarro and K. Van Acoleyen, $f(R)$ actions, cosmic acceleration and local tests of gravity, *JCAP* **0702** (2007), 022.
- [363] L. Amendola, R. Gannouji, D. Polarski, and S. Tsujikawa, Conditions for the cosmological viability of $f(R)$ dark energy models, *Phys. Rev. D* **75** (2007), 083504.
- [364] S. M. Carroll, I. Sawicki, A. Silvestri, and M. Trodden, Modified-source gravity and cosmological structure formation, *New J. Phys.* **8** (2006), 323.
- [365] Y. S. Song, W. Hu, and I. Sawicki, The large scale structure of $f(R)$ gravity, *Phys. Rev. D* **75** (2007), 044004.
- [366] R. Bean, D. Bernat, L. Pogosian, A. Silvestri, and M. Trodden, Dynamics of linear perturbations in $f(R)$ gravity, *Phys. Rev. D* **75** (2007), 064020.
- [367] T. Faulkner, M. Tegmark, E. F. Bunn, and Y. Mao, Constraining $f(R)$ gravity as a scalar tensor theory, *Phys. Rev. D* **76** (2007), 063505.
- [368] W. Hu and I. Sawicki, Models of $f(R)$ cosmic acceleration that evade solar-system tests, *Phys. Rev. D* **76** (2007), 064004.
- [369] A. A. Starobinsky, Disappearing cosmological constant in $f(R)$ gravity, *JETP Lett.* **86** (2007), 157.
- [370] S. A. Appleby and R. A. Battye, Do consistent $f(R)$ models mimic General Relativity plus Λ ?, *Phys. Lett. B* **654** (2007), 7.
- [371] S. Tsujikawa, Observational signatures of $f(R)$ dark energy models that satisfy cosmological and local gravity constraints, *Phys. Rev. D* **77** (2008), 023507.
- [372] V. Müller, H. J. Schmidt, and A. A. Starobinsky, The stability of the de-Sitter spacetime in fourth order gravity, *Phys. Lett. B* **202** (1988), 198.
- [373] V. Faraoni, de Sitter attractors in generalized gravity, *Phys. Rev. D* **70** (2004), 044037.
- [374] B. Boisseau, G. Esposito-Farese, D. Polarski, and A. A. Starobinsky, Reconstruction of a scalar-tensor theory of gravity in an accelerating universe, *Phys. Rev. Lett.* **85** (2000), 2236.

- [375] V. Miranda, S. E. Joras, I. Waga, and M. Quartin, Viable singularity-free $f(R)$ gravity without a cosmological constant, *Phys. Rev. Lett.* **102** (2009), 221101.
- [376] B. Li and J. D. Barrow, The cosmology of $f(R)$ gravity in the metric variational approach, *Phys. Rev. D* **75** (2007), 084010.
- [377] R. Gannouji, B. Moraes, and D. Polarski, The growth of matter perturbations in $f(R)$ models, *Phys. Rev. D* **0902** (2009), 034.
- [378] S. Capozziello and S. Tsujikawa, Solar system and equivalence principle constraints on $f(R)$ gravity by chameleon approach, *Phys. Rev. D* **77** (2008), 107501.
- [379] P. Brax, C. van de Bruck, A. C. Davis, and D. J. Shaw, $f(R)$ gravity and chameleon theories, *Phys. Rev. D* **78** (2008), 104021.
- [380] N. Deruelle, M. Sasaki, and Y. Sendouda, ‘Detuned’ $f(R)$ gravity and dark energy, *Phys. Rev. D* **77** (2008), 124024.
- [381] A. Dev, D. Jain, S. Jhingan, S. Nojiri, M. Sami, and I. Thongkool, Delicate $f(R)$ gravity models with disappearing cosmological constant and observational constraints on the model parameters, *Phys. Rev. D* **78** (2008), 083515.
- [382] A. V. Frolov, A singularity problem with $f(R)$ dark energy, *Phys. Rev. Lett.* **101** (2008), 061103.
- [383] T. Kobayashi and K. i. Maeda, Relativistic stars in $f(R)$ gravity, and absence thereof, *Phys. Rev. D* **78** (2008), 064019.
- [384] E. Babichev and D. Langlois, Relativistic stars in $f(R)$ gravity, *Phys. Rev. D* **80** (2009), 121501.
- [385] A. Upadhye and W. Hu, The existence of relativistic stars in $f(R)$ gravity, *Phys. Rev. D* **80** (2009), 064002.
- [386] D. N. Vollick, Curvature corrections as the source of the cosmological acceleration, *Phys. Rev. D* **68** (2003), 063510.
- [387] D. N. Vollick, On the viability of the Palatini form of $1/R$ gravity, *Class. Quant. Grav.* **21** (2004), 3813.
- [388] X. Meng and P. Wang, Modified Friedmann equations in R^{-1} -modified gravity, *Class. Quant. Grav.* **20** (2003), 4949.
- [389] E. E. Flanagan, The conformal frame freedom in theories of gravitation, *Class. Quant. Grav.* **21** (2004), 3817.
- [390] T. P. Sotiriou, Constraining $f(R)$ gravity in the Palatini formalism, *Class. Quant. Grav.* **23** (2006), 1253.
- [391] T. P. Sotiriou, $f(R)$ gravity and scalar-tensor theory, *Class. Quant. Grav.* **23** (2006), 5117.
- [392] S. Fay, R. Tavakol, and S. Tsujikawa, $f(R)$ gravity theories in Palatini formalism: Cosmological dynamics and observational constraints, *Phys. Rev. D* **75** (2007), 063509.
- [393] M. Amarzguoui, O. Elgaroy, D. F. Mota, and T. Multamaki, Cosmological constraints on $f(R)$ gravity theories within the Palatini approach, *Astron. Astrophys.* **454** (2006), 707.
- [394] T. Koivisto and H. Kurki-Suonio, Cosmological perturbations in the Palatini formulation of modified gravity, *Class. Quant. Grav.* **23** (2006), 2355.
- [395] S. Tsujikawa, K. Uddin, and R. Tavakol, Density perturbations in $f(R)$ gravity theories in metric and Palatini formalisms, *Phys. Rev. D* **77** (2008), 043007.
- [396] E. E. Flanagan, Palatini form of $1/R$ gravity, *Phys. Rev. Lett.* **92** (2004), 071101.
- [397] A. Iglesias, N. Kaloper, A. Padilla, and M. Park, How (Not) to Palatini, *Phys. Rev. D* **76** (2007), 104001.

- [398] E. Barausse, T. P. Sotiriou, and J. C. Miller, A no-go theorem for polytropic spheres in Palatini $f(R)$ gravity, *Class. Quant. Grav.* **25** (2008), 062001.
- [399] T. P. Sotiriou and V. Faraoni, $f(R)$ theories of gravity, arXiv:gr-qc/0805.1726 (2008).
- [400] Y. Fujii and K. Maeda, *The Scalar-Tensor Theory of Gravitation*, Cambridge University Press (2003).
- [401] G. Esposito-Farese and D. Polarski, Scalar-tensor gravity in an accelerating universe, *Phys. Rev. D* **63** (2001), 063504.
- [402] D. F. Torres, Quintessence, super-quintessence and observable quantities in Brans-Dicke and non-minimally coupled theories, *Phys. Rev. D* **66** (2002), 043522.
- [403] S. Nesseris and L. Perivolaropoulos, Evolving Newton's constant, extended gravity theories and SnIa data analysis, *Phys. Rev. D* **73** (2006), 103511.
- [404] T. Damour and K. Nordtvedt, Tensor-scalar cosmological models and their relaxation toward general relativity, *Phys. Rev. D* **48** (1993), 3436.
- [405] J. D. Bekenstein, Relativistic gravitation theory for the MOND paradigm, *Phys. Rev. D* **70** (2004), 083509.
- [406] S. M. Carroll, A. De Felice, V. Duvvuri, D. A. Easson, M. Trodden, and M. S. Turner, The cosmology of generalized modified gravity models, *Phys. Rev. D* **71** (2005), 063513.
- [407] A. De Felice, M. Hindmarsh, and M. Trodden, Ghosts, instabilities, and superluminal propagation in modified gravity models, *JCAP* **0608** (2006), 005.
- [408] G. Calcagni, B. de Carlos, and A. De Felice, Ghost conditions for Gauss-Bonnet cosmologies, *Nucl. Phys. B* **752** (2006), 404.
- [409] A. Nunez and S. Solganik, Ghost constraints on modified gravity, *Phys. Lett. B* **608** (2005), 189.
- [410] S. Nojiri, S. D. Odintsov, and M. Sasaki, Gauss-Bonnet dark energy, *Phys. Rev. D* **71** (2005), 123509.
- [411] G. Calcagni, S. Tsujikawa, and M. Sami, Dark energy and cosmological solutions in second-order string gravity, *Class. Quant. Grav.* **22** (2005), 3977.
- [412] B. M. N. Carter and I. P. Neupane, Towards inflation and dark energy cosmologies from modified Gauss-Bonnet theory, *JCAP* **0606** (2006), 004.
- [413] L. Amendola, C. Charmousis, and S. C. Davis, Constraints on Gauss-Bonnet gravity in dark energy cosmologies, *JCAP* **0612** (2006), 020.
- [414] T. Koivisto and D. F. Mota, Cosmology and astrophysical constraints of Gauss-Bonnet dark energy, *Phys. Lett. B* **644** (2007), 104.
- [415] T. Koivisto and D. F. Mota, Gauss-Bonnet quintessence: Background evolution, large scale structure and cosmological constraints, *Phys. Rev. D* **75** (2007), 023518.
- [416] S. Tsujikawa and M. Sami, String-inspired cosmology: Late time transition from scaling matter era to dark energy universe caused by a Gauss-Bonnet coupling, *JCAP* **0701** (2007), 006.
- [417] J. c. Hwang and H. Noh, Classical evolution and quantum generation in generalized gravity theories including string corrections and tachyon: Unified analyses, *Phys. Rev. D* **71** (2005), 063536.
- [418] Z. K. Guo, N. Ohta, and S. Tsujikawa, Realizing scale-invariant density perturbations in low-energy effective string theory, *Phys. Rev. D* **75** (2007), 023520.
- [419] L. Amendola, C. Charmousis, and S. C. Davis, Solar system constraints on Gauss-Bonnet mediated dark energy, *JCAP* **0710** (2007), 004.
- [420] S. Nojiri and S. D. Odintsov, Modified Gauss-Bonnet theory as gravitational alternative for dark energy, *Phys. Lett. B* **631** (2005), 1.
- [421] B. Li, J. D. Barrow, and D. F. Mota, The cosmology of modified Gauss-Bonnet gravity, *Phys. Rev. D* **76** (2007), 044027.

- [422] A. De Felice and S. Tsujikawa, Construction of cosmologically viable $f(\mathcal{G})$ gravity models, *Phys. Lett. B* **675** (2009), 1.
- [423] S. Y. Zhou, E. J. Copeland, and P. M. Saffin, Cosmological constraints on $f(\mathcal{G})$ dark energy models, *JCAP* **0907** (2009), 009.
- [424] A. De Felice and S. Tsujikawa, Solar system constraints on $f(\mathcal{G})$ gravity models, *Phys. Rev. D* **80** (2009), 063516.
- [425] A. De Felice, D. F. Mota, and S. Tsujikawa, Matter instabilities in general Gauss-Bonnet gravity, *Phys. Rev. D* **81** (2010), 023532.
- [426] L. Randall and R. Sundrum, A large mass hierarchy from a small extra dimension, *Phys. Rev. Lett.* **83** (1999), 3370.
- [427] L. Randall and R. Sundrum, An alternative to compactification, *Phys. Rev. Lett.* **83** (1999), 4690.
- [428] P. Binetruy, C. Deffayet, and D. Langlois, Non-conventional cosmology from a brane-universe, *Nucl. Phys. B* **565** (2000), 269.
- [429] P. Binetruy, C. Deffayet, U. Ellwanger, and D. Langlois, Brane cosmological evolution in a bulk with cosmological constant, *Phys. Lett. B* **477** (2000), 285.
- [430] T. Shiromizu, K. i. Maeda, and M. Sasaki, The Einstein equations on the 3-brane world, *Phys. Rev. D* **62** (2000), 024012.
- [431] G. R. Dvali, G. Gabadadze, and M. Porrati, 4D gravity on a brane in 5D Minkowski space, *Phys. Lett. B* **485** (2000), 208.
- [432] G. R. Dvali and G. Gabadadze, Gravity on a brane in infinite-volume extra space, *Phys. Rev. D* **63** (2001), 065007.
- [433] C. Deffayet, Cosmology on a brane in Minkowski bulk, *Phys. Lett. B* **502** (2001), 199.
- [434] C. Deffayet, G. R. Dvali, and G. Gabadadze, Accelerated universe from gravity leaking to extra dimensions, *Phys. Rev. D* **65** (2002), 044023.
- [435] I. Sawicki and S. M. Carroll, Cosmological structure evolution and CMB anisotropies in DGP braneworlds, arXiv:astro-ph/0510364 (2005).
- [436] M. Fairbairn and A. Goobar, Supernova limits on brane world cosmology, *Phys. Lett. B* **642** (2006), 432.
- [437] R. Maartens and E. Majerotto, Observational constraints on self-accelerating cosmology, *Phys. Rev. D* **74** (2006), 023004.
- [438] U. Alam and V. Sahni, Confronting braneworld cosmology with supernova data and baryon oscillations, *Phys. Rev. D* **73** (2006), 084024.
- [439] Y. S. Song, I. Sawicki, and W. Hu, Large-scale tests of the DGP model, *Phys. Rev. D* **75** (2007), 064003.
- [440] A. Lue, R. Scoccimarro, and G. D. Starkman, Probing Newton's constant on vast scales: DGP gravity, cosmic acceleration and large scale structure, *Phys. Rev. D* **69** (2004), 124015.
- [441] K. Koyama and R. Maartens, Structure formation in the DGP cosmological model, *JCAP* **0601** (2006), 016.
- [442] R. Durrer and R. Maartens, Dark energy and modified gravity, arXiv:astro-ph/0811.4132 (2008).
- [443] D. Gorbunov, K. Koyama, and S. Sibiryakov, More on ghosts in DGP model, *Phys. Rev. D* **73** (2006), 044016.
- [444] C. Charmousis, R. Gregory, N. Kaloper, and A. Padilla, DGP spectroscopy, *JHEP* **0610** (2006), 066.
- [445] K. Tomita, Distances and lensing in cosmological void models, *Astrophys. J.* **529** (2000), 38.
- [446] K. Tomita, A local void and the accelerating universe, *Mon. Not. Roy. Astron. Soc.* **326** (2001), 287.

- [447] M. N. Celerier, Do we really see a cosmological constant in the supernovae data?, *Astron. Astrophys.* **353** (2000), 63.
- [448] H. Iguchi, T. Nakamura, and K. i. Nakao, Is dark energy the only solution to the apparent acceleration of the present universe?, *Prog. Theor. Phys.* **108** (2002), 809.
- [449] H. Alnes, M. Amarzguoui, and O. Gron, An inhomogeneous alternative to dark energy?, *Phys. Rev. D* **73** (2006), 083519.
- [450] A. E. Romano and M. Sasaki, On the relation between positive averaged acceleration and physical observables in LTB spaces, arXiv:0905.3342 [astro-ph.CO] (2009).
- [451] N. Mustapha, C. Hellaby, and G. F. R. Ellis, Large scale inhomogeneity versus source evolution: Can we distinguish them observationally?, *Mon. Not. Roy. Astron. Soc.* **292** (1997), 817.
- [452] J. Garcia-Bellido and T. Haugboelle, Confronting Lemaître-Tolman-Bondi models with observational cosmology, *JCAP* **0804** (2008), 003.
- [453] K. Enqvist, Lemaître-Tolman-Bondi model and accelerating expansion, *Gen. Rel. Grav.* **40** (2008), 451.
- [454] J. Kristian and R. K. Sachs, Observations in cosmology, *Astrophys. J.* **143** (1966), 379.
- [455] H. Alnes and M. Amarzguoui, CMB anisotropies seen by an off-center observer in a spherically symmetric inhomogeneous universe, *Phys. Rev. D* **74** (2006), 103520.
- [456] M. Cruz, E. Martinez-Gonzalez, P. Vielva, J. M. Diego, M. Hobson, and N. Turok, The CMB cold spot: texture, cluster or void?, arXiv:astro-ph/0804.2904 (2008).
- [457] I. Masina and A. Notari, The cold spot as a large void: Rees-Sciama effect on CMB power spectrum and bispectrum, *JCAP* **0902** (2009), 019.
- [458] J. Garcia-Bellido and T. Haugboelle, Looking the void in the eyes and the kinematic Sunyaev Zeldovich effect in Lemaître Tolman Bondi models, *JCAP* **0809** (2008), 016.
- [459] R. R. Caldwell and A. Stebbins, A test of the Copernican Principle, *Phys. Rev. Lett.* **100** (2008), 191302.
- [460] F. Occhionero, P. Santangelo, and N. Vittorio, Holes in cosmology, *Astron. Astrophys.* **117** (1983), 365.
- [461] T. Biswas, R. Mansouri, and A. Notari, Non-linear structure formation and apparent acceleration: an Investigation, *JCAP* **0712** (2007), 017.
- [462] C. Clarkson, B. Bassett, and T. C. Lu, A general test of the Copernican Principle, *Phys. Rev. Lett.* **101** (2008), 011301.
- [463] M. H. Partovi and B. Mashhoon, Toward verification of large-scale homogeneity in cosmology, *Astrophys. J.* **276** (1984), 4.
- [464] C. Quercellini, M. Quartin, and L. Amendola, Possibility of detecting anisotropic expansion of the Universe by very accurate astrometry measurements, *Phys. Rev. Lett.* **102** (2009), 151302.
- [465] S. Rasanen, Dark energy from backreaction, *JCAP* **0402** (2004), 003.
- [466] E. W. Kolb, S. Matarrese, A. Notari, and A. Riotto, The effect of inhomogeneities on the expansion rate of the universe, *Phys. Rev. D* **71** (2005), 023524.
- [467] E. W. Kolb, S. Matarrese, and A. Riotto, On cosmic acceleration without dark energy, *New J. Phys.* **8** (2006), 322.
- [468] C. M. Hirata and U. Seljak, Can superhorizon cosmological perturbations explain the acceleration of the universe?, *Phys. Rev. D* **72** (2005), 083501.
- [469] P. Martineau and R. H. Brandenberger, The effects of gravitational back-reaction on cosmological perturbations, *Phys. Rev. D* **72** (2005), 023507.

- [470] A. Ishibashi and R. M. Wald, Can the acceleration of our universe be explained by the effects of inhomogeneities?, *Class. Quant. Grav.* **23** (2006), 235.
- [471] M. Kasai, H. Asada, and T. Futamase, Toward a no-go theorem for accelerating universe by non-linear backreaction, *Prog. Theor. Phys.* **115** (2006), 827.
- [472] T. Buchert, Dark energy from structure – A status report, *Gen. Rel. Grav.* **40** (2008), 467.
- [473] S. Rasanen, The effect of structure formation on the expansion of the universe, *Int. J. Mod. Phys. D* **17** (2009), 2543.
- [474] E. W. Kolb, V. Marra, and S. Matarrese, Cosmological background solutions and cosmological backreactions, arXiv:astro-ph/0901.4566 (2009).
- [475] F. Piazza, Modifying gravity in the infra-red by imposing an ‘ultra-strong’ equivalence principle, arXiv:hep-th/0904.4299 (2009).
- [476] F. Piazza, The IR-completion of gravity: What happens at Hubble scales?, *New J. Phys.* **11** (2009), 113050.
- [477] V. Silveira and I. Waga, Decaying Lambda cosmologies and power spectrum, *Phys. Rev. D* **50** (1994), 4890.
- [478] L. M. Wang and P. J. Steinhardt, Cluster abundance constraints on quintessence models, *Astrophys. J.* **508** (1998), 483.
- [479] Y. Gong, M. Ishak, and A. Wang, Growth factor parametrization in curved space, *Phys. Rev. D* **80** (2009), 023002.
- [480] E. V. Linder, Cosmic growth history and expansion history, *Phys. Rev. D* **72** (2005), 043529.
- [481] C. Di Porto and L. Amendola, Observational constraints on the linear fluctuation growth rate, *Phys. Rev. D* **77** (2008), 083508.
- [482] N. Bartolo, P. S. Corasaniti, A. R. Liddle, and M. Malquarti, Perturbations in cosmologies with a scalar field and a perfect fluid, *Phys. Rev. D* **70** (2004), 043532.
- [483] L. R. Abramo, F. Finelli, and T. S. Pereira, Constraining Born-Infeld models of dark energy with CMB anisotropies, *Phys. Rev. D* **70** (2004), 063517.
- [484] L. Amendola, Phantom energy mediates a long-range repulsive force, *Phys. Rev. Lett.* **93** (2004), 181102.
- [485] S. Tsujikawa, Reconstruction of general scalar-field dark energy models, *Phys. Rev. D* **72** (2005), 083152.
- [486] L. Amendola and D. Tocchini-Valentini, Baryon bias and structure formation in an accelerating universe, *Phys. Rev. D* **66** (2002), 043528.
- [487] L. Amendola and C. Quercellini, Skewness as a test of the equivalence principle, *Phys. Rev. Lett.* **92** (2004), 181102.
- [488] W. Hu, R. Barkana, and A. Gruzinov, Cold and fuzzy dark matter, *Phys. Rev. Lett.* **85** (2000), 1158.
- [489] L. Amendola and R. Barbieri, Dark matter from an ultra-light pseudo-Goldstone-boson, *Phys. Lett. B* **642** (2006), 192.
- [490] Y. S. Song, H. Peiris, and W. Hu, Cosmological constraints on $f(R)$ acceleration models, *Phys. Rev. D* **76** (2007), 063517.
- [491] S. Tsujikawa and T. Tatekawa, The effect of modified gravity on weak lensing, *Phys. Lett. B* **665** (2008), 325.
- [492] F. Schmidt, Weak lensing probes of modified gravity, *Phys. Rev. D* **78** (2008), 043002.
- [493] P. Macdonald *et al.*, The linear theory power spectrum from the Lyman-alpha forest in the Sloan Digital Sky Survey, *Astrophys. J.* **635** (2005), 761.
- [494] M. Viel and M. G. Haehnelt, Cosmological and astrophysical parameters from the SDSS flux power spectrum and hydrodynamical simulations of the Lyman-alpha forest, *Mon. Not. Roy. Astron. Soc.* **365** (2006), 231.

- [495] K. Koyama and F. P. Silva, Non-linear interactions in a cosmological background in the DGP braneworld, *Phys. Rev. D* **75** (2007), 084040.
- [496] K. Yamamoto, D. Parkinson, T. Hamana, R. C. Nichol, and Y. Suto, Optimizing future imaging survey of galaxies to confront dark energy and modified gravity models, *Phys. Rev. D* **76** (2007), 23504.
- [497] L. Amendola, M. Kunz, and D. Sapone, Measuring the dark side (with weak lensing), *JCAP* **0804** (2008), 013.
- [498] S. Tsujikawa, Matter density perturbations and effective gravitational constant in modified gravity models of dark energy, *Phys. Rev. D* **76** (2007), 023514.
- [499] T. Tatekawa, Lagrangian perturbation theory in Newtonian cosmology, *Recent Res. Devel. Phys.* **2** (2005), 1 [arXiv:astro-ph/0412025].
- [500] F. Bernardeau, Skewness and Kurtosis in large scale cosmic fields, *Astrophys. J.* **433** (1994), 1.
- [501] B. Jain and E. Bertschinger, Second order power spectrum and non-linear evolution at high redshift, *Astrophys. J.* **431** (1994), 495.
- [502] J. N. Fry, The Galaxy correlation hierarchy in perturbation theory, *Astrophys. J.* **279** (1984), 499.
- [503] M. H. Goroff, B. Grinstein, S. J. Rey, and M. B. Wise, Coupling of modes of cosmological mass density fluctuations, *Astrophys. J.* **311** (1986), 6.
- [504] P. Catelan, F. Lucchin, S. Matarrese, and L. Moscardini, Eulerian perturbation theory in nonflat universes: second order approximation, *Mon. Not. Roy. Astron. Soc.* **276** (1995), 39.
- [505] F. Bernardeau, S. Colombi, E. Gaztanaga, and R. Scoccimarro, Large-scale structure of the universe and cosmological perturbation theory, *Phys. Rept.* **367** (2002), 1.
- [506] M. Kamionkowski and A. Buchalter, Weakly non-linear clustering for arbitrary expansion histories, *Astrophys. J.* **514** (1999), 7.
- [507] E. Gaztanaga and J. A. Lobo, Non-linear gravitational growth of large scale structures inside and outside standard cosmology, *Astrophys. J.* **548** (2001), 47.
- [508] R. Juszkiewicz, F. R. Bouchet, and S. Colombi, Skewness induced by gravity, *Astrophys. J.* **412** (1993), L9.
- [509] K. Benabed and F. Bernardeau, Testing quintessence models with large-scale structure growth, *Phys. Rev. D* **64** (2001), 083501.
- [510] T. Tatekawa and S. Tsujikawa, Second-order matter density perturbations and skewness in scalar-tensor modified gravity models, *JCAP* **0809** (2008), 009.
- [511] S. Matarrese and M. Pietroni, Resumming cosmic perturbations, *JCAP* **0706** (2007), 026.
- [512] M. Crocce and R. Scoccimarro, Renormalized cosmological perturbation theory, *Phys. Rev. D* **73** (2006), 063519.
- [513] T. Matsubara, Nonlinear perturbation theory with halo bias and redshift-space distortions via the Lagrangian picture, *Phys. Rev. D* **78** (2008), 083519.
- [514] S. Matarrese, L. Verde, and A. F. Heavens, Large-scale bias in the universe: bispectrum method, *Mon. Not. Roy. Astron. Soc.* **290** (1997), 651.
- [515] L. Verde, A. F. Heavens, S. Matarrese, and L. Moscardini, Large-scale bias in the universe II: redshift space bispectrum, *Mon. Not. Roy. Astron. Soc.* **300** (1998), 747.
- [516] P. Fosalba and E. Gaztanaga, Cosmological perturbation theory and the spherical collapse model: Part I. Gaussian initial conditions, *Mon. Not. Roy. Astron. Soc.* **301** (1998), 503.
- [517] W. H. Press and P. Schechter, Formation of galaxies and clusters of galaxies by selfsimilar gravitational condensation, *Astrophys. J.* **187** (1974), 425.
- [518] J. A. Peacock, *Cosmological Physics*, Cambridge University Press (1999).

- [519] N. N. Weinberg and M. Kamionkowski, Constraining dark energy from the abundance of weak gravitational lenses, *Mon. Not. Roy. Astron. Soc.* **341** (2003), 251.
- [520] D. F. Mota and C. van de Bruck, On the spherical collapse model in dark energy cosmologies, *Astron. Astrophys.* **421** (2004), 71.
- [521] I. Maor and O. Lahav, On virialization with dark energy, *JCAP* **0507** (2005), 003.
- [522] L. R. Abramo, R. C. Batista, L. Liberato, and R. Rosenfeld, Physical approximations for the non-linear evolution of perturbations in inhomogeneous dark energy scenarios, *Phys. Rev. D* **79** (2009), 023516.
- [523] R. Mainini and S. Bonometto, Mass functions in coupled dark energy models, *Phys. Rev. D* **74** (2006), 043504.
- [524] J. R. Bond, S. Cole, G. Efstathiou, and N. Kaiser, Excursion set mass functions for hierarchical Gaussian fluctuations, *Astrophys. J.* **379** (1991), 440.
- [525] R. K. Sheth and G. Tormen, Large scale bias and the peak background split, *Mon. Not. Roy. Astron. Soc.* **308** (1999), 119.
- [526] S. Sasaki, Formation rate of bound objects in the hierarchical clustering model, *Pub. Astr. Soc. Japan* **46** (1994), 427.
- [527] A. Jenkins *et al.*, Mass function of dark matter halos, *Mon. Not. Roy. Astron. Soc.* **321** (2001), 372.
- [528] P. Rosati, S. Borgani, and C. Norman, The evolution of X-ray clusters of galaxies, *Ann. Rev. Astron. Astrophys.* **40** (2002), 539.
- [529] J. L. Tinker *et al.*, Toward a halo mass function for precision cosmology: the limits of universality, *Astrophys. J.* **688** (2008), 709.
- [530] A. J. S. Hamilton, A. Matthews, P. Kumar, and E. Lu, Reconstructing the primordial spectrum of fluctuations of the universe from the observed nonlinear clustering of galaxies, *Astrophys. J.* **374** (1991), L1.
- [531] J. A. Peacock and S. J. Dodds, Non-linear evolution of cosmological power spectra, *Mon. Not. Roy. Astron. Soc.* **280** (1996), L19.
- [532] R. E. Smith *et al.*, The Virgo Consortium Collaboration, Stable clustering, the halo model and non-linear cosmological power spectra, *Mon. Not. Roy. Astron. Soc.* **341** (2003), 1311.
- [533] A. Klypin, A. V. Maccio, R. Mainini, and S. A. Bonometto, Halo properties in models with dynamical dark energy, *Astrophys. J.* **599** (2003), 31.
- [534] P. McDonald, H. Trac, and C. Contaldi, Dependence of the non-linear mass power spectrum on the equation of state of dark energy, *Mon. Not. Roy. Astron. Soc.* **366** (2006), 547.
- [535] L. Casarini, A. V. Maccio, and S. A. Bonometto, Dynamical dark energy simulations: high accuracy power spectra at high redshift, *JCAP* **0903** (2009), 014.
- [536] A. V. Maccio, C. Quercellini, R. Mainini, L. Amendola, and S. A. Bonometto, N-body simulations for coupled dark energy: halo mass function and density profiles, *Phys. Rev. D* **69** (2004), 123516.
- [537] M. Baldi, V. Pettorino, G. Robbers, and V. Springel, N-body simulations of coupled dark energy cosmologies, arXiv:astro-ph/0812.3901 (2008).
- [538] I. Laszlo and R. Bean, Non-linear growth in modified gravity theories of dark energy, *Phys. Rev. D* **77** (2008), 024048.
- [539] H. Oyaizu, Non-linear evolution of $f(R)$ cosmologies I: methodology, *Phys. Rev. D* **78** (2008), 123523.
- [540] H. Oyaizu, M. Lima, and W. Hu, Non-linear evolution of $f(R)$ cosmologies II: power spectrum, *Phys. Rev. D* **78** (2008), 123524.

- [541] K. Koyama, A. Taruya, and T. Hiramatsu, Non-linear evolution of matter power spectrum in modified theory of gravity, *Phys. Rev. D* **79** (2009), 123512.
- [542] B. Li and H. Zhao, Structure formation by a fifth force: N-body versus linear simulations, *Phys. Rev. D* **80** (2009), 044027.
- [543] R. Kass and A. Raftery, Bayes factors and model uncertainty, *J. Amer. Statist. Assoc.* **90** (1995), 773.
- [544] R. Trotta, Bayes in the sky: Bayesian inference and model selection in cosmology, *Contemp. Phys.* **49** (2008), 71.
- [545] E. V. Linder and R. Miquel, Cosmological model selection: Statistics and physics, *Int. J. Mod. Phys. D* **17** (2008), 2315.
- [546] A. R. Liddle, P. Stefano Corasaniti, M. Kunz, P. Mukherjee, D. Parkinson, and R. Trotta, Comment on ‘Tainted evidence: cosmological model selection versus fitting, by Eric V. Linder and Ramon Miquel (astro-ph/0702542v2)’, arXiv:astro-ph/0703285 (2007).
- [547] A. Albrecht *et al.*, Report of the dark energy task force, arXiv:astro-ph/0609591 (2006).
- [548] M. Tegmark, Measuring cosmological parameters with galaxy surveys, *Phys. Rev. Lett.* **79** (1997), 3806.
- [549] H. J. Seo and D. J. Eisenstein, Probing dark energy with baryonic acoustic oscillations from future large galaxy redshift surveys, *Astrophys. J.* **598** (2003), 720.
- [550] M. Tegmark, A. Taylor, and A. Heavens, Karhunen-Loeve eigenvalue problems in cosmology: how should we tackle large data sets?, *Astrophys. J.* **480** (1997), 22.
- [551] D. J. Eisenstein, W. Hu, and M. Tegmark, Cosmic complementarity: Joint parameter estimation from CMB experiments and redshift surveys, *Astrophys. J.* **518** (1999), 2.
- [552] D. Huterer and G. Starkman, Parameterization of dark-energy properties: A principal-component approach, *Phys. Rev. Lett.* **90** (2003), 031301.
- [553] A. J. Albrecht and G. Bernstein, Evaluating dark energy probes using multidimensional dark energy parameters, *Phys. Rev. D* **75** (2007), 103003.
- [554] A. J. Albrecht *et al.*, Findings of the joint dark energy mission figure of Merit Science Working Group, arXiv:astro-ph/0901.0721 (2009).
- [555] R. G. Crittenden and N. Turok, Looking for Λ with the Rees-Sciama effect, *Phys. Rev. Lett.* **76** (1996), 575.
- [556] J. D. Jackson, *Classical Electrodynamics*, Wiley and Sons, New York, 3rd edition (2002).
- [557] R. Bean and O. Dore, Probing dark energy perturbations: the dark energy equation of state and speed of sound as measured by WMAP, *Phys. Rev. D* **69** (2004), 083503.
- [558] J. Garriga, L. Pogosian, and T. Vachaspati, Forecasting cosmic doomsday from CMB/LSS cross-correlations, *Phys. Rev. D* **69** (2004), 063511.
- [559] A. Cooray, Integrated Sachs-Wolfe effect: Large scale structure correlation, *Phys. Rev. D* **65** (2002), 103510.
- [560] J. Garriga, L. Pogosian, and T. Vachaspati, Forecasting cosmic doomsday from CMB/LSS cross-correlations, *Phys. Rev. D* **69** (2004), 063511.
- [561] T. Giannantonio, R. Scranton, R. G. Crittenden, R. C. Nichol, S. P. Boughn, A. D. Myers, and G. T. Richards, Combined analysis of the integrated Sachs-Wolfe effect and cosmological implications, *Phys. Rev. D* **77** (2008), 123520.
- [562] S. Ho, C. Hirata, N. Padmanabhan, U. Seljak, and N. Bahcall, Correlation of CMB with large-scale structure: I. ISW Tomography and cosmological implications, *Phys. Rev. D* **78** (2008), 043519.

- [563] M. Douspis, P. G. Castro, C. Caprini, and N. Aghanim, Optimising large galaxy surveys for ISW detection, arXiv:astro-ph/0802.0983 (2008).
- [564] J. Weller and A. M. Lewis, Large scale cosmic microwave background anisotropies and dark energy, *Mon. Not. Roy. Astron. Soc.* **346** (2003), 987.
- [565] H. J. Seo and D. J. Eisenstein, Improved forecasts for the baryon acoustic oscillations and cosmological distance scale, *Astrophys. J.* **665** (2007), 14.
- [566] C. Alcock and B. Paczynski, An evolution free test for non-zero cosmological constant, *Nature* **281** (1979), 358.
- [567] W. Balliger, J. A. Peacock, and A. F. Heavens, Measuring the cosmological constant with redshift surveys, *Mon. Not. Roy. Astron. Soc.* **282** (1996), 877.
- [568] A. Rassat *et al.*, Deconstructing baryon acoustic oscillations: A comparison of methods, arXiv:astro-ph/0810.0003 (2008).
- [569] L. Amendola, C. Quercellini, and E. Giallongo, Constraints on perfect fluid and scalar field dark energy models from future redshift surveys, *Mon. Not. Roy. Astron. Soc.* **357** (2005), 429.
- [570] C. Blake, D. Parkinson, B. Bassett, K. Glazebrook, M. Kunz, and R. C. Nichol, Universal fitting formulae for baryon oscillation surveys, *Mon. Not. Roy. Astron. Soc.* **365** (2006), 255.
- [571] D. Dolney, B. Jain, and M. Takada, Baryon oscillations and dark-energy constraints from imaging surveys, *Mon. Not. Roy. Astron. Soc.* **366** (2006), 884.
- [572] E. Hawkins *et al.*, The 2dF Galaxy Redshift Survey: correlation functions, peculiar velocities and the matter density of the universe, *Mon. Not. Roy. Astron. Soc.* **346** (2003), 78.
- [573] N. P. Ross *et al.*, The 2dF-SDSS LRG and QSO Survey: the LRG 2-point correlation function and redshift-space distortions, *Mon. Not. Roy. Astron. Soc.* **381** (2007), 573.
- [574] L. Guzzo *et al.*, A test of the nature of cosmic acceleration using galaxy redshift distortions, *Nature* **451** (2008), 541.
- [575] A. J. S. Hamilton, Measuring Ω and the real correlation function from the redshift correlation function, *Astrophys. J.* **385** (1992), L5.
- [576] W. Saunders, M. Rowan-Robinson, and A. Lawrence, The spatial correlation function of IRAS galaxies on small and intermediate scales, *Mon. Not. Roy. Astron. Soc.* **258** (1992), 134.
- [577] M. Bartelmann and P. Schneider, Weak gravitational lensing, *Phys. Rept.* **340** (2001), 291.
- [578] P. Schneider, *Weak Gravitational Lensing, Gravitational Lensing: Strong, Weak and Micro*, Saas-Fee Advanced Courses, Volume 33. ISBN 978-3-540-30309-1, Springer-Verlag, Berlin Heidelberg (2006), 269.
- [579] D. J. Bacon, A. R. Refregier, and R. S. Ellis, Detection of weak gravitational lensing by large-scale structure, *Mon. Not. Roy. Astron. Soc.* **318** (2000), 625.
- [580] N. Kaiser, G. Wilson, and G. A. Luppino, Large-scale cosmic shear measurements, arXiv:astro-ph/0003338 (2000).
- [581] L. van Waerbeke *et al.*, Detection of correlated galaxy ellipticities on CFHT data: first evidence for gravitational lensing by large-scale structures, *Astron. Astrophys.* **358** (2000), 30.
- [582] D. M. Wittman, J. A. Tyson, D. Kirkman, I. Dell'Antonio, and G. Bernstein, Detection of weak gravitational lensing distortions of distant galaxies by cosmic dark matter at large scales, *Nature* **405** (2000), 143.
- [583] H. Hoekstra and B. Jain, Weak gravitational lensing and its cosmological applications, *ARNPS* **58** (2008), 99.

- [584] A. Amara and A. Refregier, Optimal surveys for weak lensing tomography, *Mon. Not. Roy. Astron. Soc.* **381** (2007), 1018.
- [585] N. Kaiser, Weak gravitational lensing of distant galaxies, *Astrophys. J.* **388** (1992), 272.
- [586] J. Miralda-Escude, The correlation function of galaxy ellipticities produced by gravitational lensing, *Astrophys. J.* **380** (1991), 1.
- [587] W. Hu, Power spectrum tomography with weak lensing, *Astrophys. J.* **522** (1999), L21.
- [588] C. M. Hirata and U. Seljak, Intrinsic alignment-lensing interference as a contaminant of cosmic shear, *Phys. Rev. D* **70** (2004), 063526.
- [589] W. Hu and M. Tegmark, Weak lensing: Prospects for measuring cosmological parameters, *Astrophys. J.* **514** (1999), L65.
- [590] M. Takada and B. Jain, Cosmological parameters from lensing power spectrum and bispectrum tomography, *Mon. Not. Roy. Astron. Soc.* **348** (2004), 897.
- [591] G. M. Bernstein, Comprehensive two-point analyses of weak gravitational lensing surveys, *Astrophys. J.* **695** (2009), 652.
- [592] L. Fu *et al.*, Very weak lensing in the CFHTLS Wide: Cosmology from cosmic shear in the linear regime, *Astron. Astrophys.* **479** (2008), 9.
- [593] S. Borgani, Cosmology with clusters of galaxies, arXiv:astro-ph/0605575 (2006).
- [594] S. L. Shapiro and S. A. Teukolsky, *Black holes, White Dwarfs, and Neutron Stars*, Wiley, New York (1983).
- [595] A. Cavaliere and R. Fusco-Femiano, X-rays from hot plasma in clusters of galaxies, *Astron. Astrophys.* **49** (1976), 137.
- [596] A. Vikhlinin, A. Kravtsov, W. Forman, C. Jones, M. Markevitch, S. S. Murray, and L. Van Speybroeck, Chandra sample of nearby relaxed galaxy clusters: mass, gas fraction, and mass-temperature relation, *Astrophys. J.* **640** (2006), 691.
- [597] K. Pedersen and H. Dahle, Calibration of the mass-temperature relation for clusters of galaxies using weak gravitational lensing, arXiv:astro-ph/0603260 (2006).
- [598] S. Sasaki, A new method to estimate cosmological parameters using the baryon fraction in clusters of galaxies, *Pub. Astr. Soc. Japan* **48** (1996), L119.
- [599] U.-L. Pen, Measuring the universal deceleration using angular diameter distances to clusters of galaxies, *New Astron.*, **2** (1997), 309.
- [600] S. Ettori, K. Dolag, S. Borgani, and G. Murante, The baryon fraction in hydrodynamical simulations of galaxy clusters, *Mon. Not. Roy. Astron. Soc.* **365** (2006), 1021.
- [601] S. W. Allen, D. A. Rapetti, R. W. Schmidt, H. Ebeling, G. Morris, and A. C. Fabian, Improved constraints on dark energy from Chandra X-ray observations of the largest relaxed galaxy clusters, *Mon. Not. Roy. Astron. Soc.* **383** (2008), 879.
- [602] S. Ettori *et al.*, The cluster gas mass fraction as a cosmological probe: a revised study, arXiv:astro-ph/0904.2740 (2009).
- [603] D. Rapetti, S. W. Allen, A. Mantz, and H. Ebeling, Constraints on modified gravity from the observed X-ray luminosity function of galaxy clusters, arXiv:astro-ph/0812.2259 (2008).
- [604] N. R. Tanvir *et al.*, A glimpse of the end of the dark ages: the gamma-ray burst of 23 April 2009 at redshift 8.3, arXiv:astro-ph/0906.1577 (2009).
- [605] G. Ghirlanda, G. Ghisellini, and C. Firmani, Gamma ray bursts as standard candles to constrain the cosmological parameters, *New J. Phys.* **8** (2006), 123.
- [606] L. Amati, C. Guidorzi, F. Frontera, M. Della Valle, F. Finelli, R. Landi, and E. Montanari, Measuring the cosmological parameters with the $E_{p,i}$ - E_{iso} correlation of gamma-ray bursts, *Mon. Not. Roy. Astron. Soc.* **391** (2008), 577.

- [607] A. P. Bouvier, V. Petrosian, and F. Ryde, On the use of gamma-ray bursts as a cosmological tool, American Astronomical Society, HEAD meeting, *Bull. Amer. Astron. Soc.* **38** (2006), 369.
- [608] S. Basilakos and L. Perivolaropoulos, Testing GRBs as standard candles, *Mon. Not. Roy. Astron. Soc.* **391** (2008), 411.
- [609] C. Firmani, G. Ghisellini, V. Avila-Reese, and G. Ghirlanda, Discovery of a tight correlation among the prompt emission properties of long gamma ray bursts, *Mon. Not. Roy. Astron. Soc.* **370** (2006), 185.
- [610] J. Dunlop, J. Peacock, H. Spinrad, A. Dey, R. Jimenez, D. Stern, and R. Windhorst, A 3.5-Gyr-old galaxy at redshift 1.55, *Nature* **381** (1996), 581.
- [611] J. A. S. Lima and J. S. Alcaniz, Constraining the cosmic equation of state from old galaxies at high redshift, *Mon. Not. Roy. Astron. Soc.* **317** (2000), 893.
- [612] J. Simon, L. Verde, and R. Jimenez, Constraints on the redshift dependence of the dark energy potential, *Phys. Rev. D* **71** (2005), 123001.
- [613] M. A. Dantas and J. S. Alcaniz, Current lookback time-redshift bounds on dark energy, *Phys. Lett. B* **79** (2009), 423.
- [614] P. Schneider, J. Ehlers, and E. E. Falco, *Gravitational Lenses*, Springer-Verlag, Berlin (1992).
- [615] E. L. Turner, J. P. Ostriker, and J. R. I. Gott, The statistics of gravitational lenses: The distributions of image angular separations and lens redshifts, *Astrophys. J.* **284** (1984), 1.
- [616] C. Porciani, and P. Madau, Gravitational lensing of distant supernovae in cold dark matter universes, *Astrophys. J.* **532** (2000), 679.
- [617] L. X. Li and J. P. Ostriker, Semi-analytical models for lensing by dark halos: I. Splitting angles, *Astrophys. J.* **566** (2002), 652.
- [618] J. F. Navarro, C. S. Frenk, and S. D. M. White, The structure of cold dark matter halos, *Astrophys. J.* **462** (1996), 563.
- [619] H. Zhao, Analytical models for galactic nuclei, *Mon. Not. Roy. Astron. Soc.* **278** (1996), 488.
- [620] S. T. Myers *et al.* [CLASS Collaboration], The cosmic lens all-sky survey: I. Source selection and observations, *Mon. Not. Roy. Astron. Soc.* **334** (2002), 1.
- [621] A. S. Bolton *et al.*, The Sloan Lens ACS Survey. V. The full ACS strong-lens sample, *Astrophys. J.* **682** (2008), 964.
- [622] M. Oguri *et al.*, The Sloan Digital Sky Survey Quasar Lens Search. III. Constraints on dark energy from the Third Data Release Quasar Lens Catalog, *Astron. J.* **135** (2008), 512.
- [623] M. Bartelmann, M. Meneghetti, F. Perrotta, C. Baccigalupi, and L. Moscardini, Arc statistics in cosmological models with dark energy, *Astron. Astrophys.* **409** (2003), 449.
- [624] M. Meneghetti, M. Bartelmann, K. Dolag, L. Moscardini, F. Perrotta, C. Baccigalupi, and G. Tormen, Strong lensing by cluster-sized haloes in dark-energy cosmologies, *New Astron. Rev.* **49** (2005), 111.
- [625] G. L. Li, S. Mao, Y. P. Jing, H. J. Mo, L. Gao, and W. P. Lin, The giant arc statistics in the three year WMAP cosmological model, *Mon. Not. Roy. Astron. Soc. Lett.* **372** (2006), L73.
- [626] M. Oguri, Gravitational lens time delays: A statistical assessment of lens model dependences and implications for the global Hubble constant, *Astrophys. J.* **660** (2007), 1.
- [627] Q. J. Zhang, L. M. Cheng, and Y. L. Wu, Constraining dark energy from splitting angle statistic of strong gravitational lenses, *Astrophys. J.* **694** (2009), 1402.

- [628] A. Sandage, The change of redshift and apparent luminosity of galaxies due to the deceleration of selected expanding universes, *Astrophys. J.* **136** (1962), 319.
- [629] M. M. Davis and L. S. May, New observations of the radio absorption line in 3C 286, with potential application to the direct measurement of cosmological deceleration, *Astrophys. J.* **219** (1978), 1.
- [630] K. Lake, Comment on the time evolution of the cosmological redshift, *Astrophys. J.* **247** (1981), 17.
- [631] R. Rudiger, Time variation of the cosmological redshift in Dicke-Brans-Jordan cosmologies, *Astrophys. J.* **260** (1982), 33.
- [632] A. Loeb, Direct measurement of cosmological parameters from the cosmic deceleration of extragalactic objects, arXiv:astro-ph/9802122 (1998).
- [633] J. Liske *et al.*, Cosmic dynamics in the era of extremely large telescopes, *Mon. Not. Roy. Astron. Soc. Lett.* **386** (2008), 1192.
- [634] P. S. Corasaniti, D. Huterer, and A. Melchiorri, Exploring the dark energy redshift desert with the Sandage-Loeb test, *Phys. Rev. D* **75** (2007), 062001.
- [635] A. Balbi and C. Quercellini, The time evolution of cosmological redshift as a test of dark energy, *Mon. Not. Roy. Astron. Soc.* **382** (2007), 1623.
- [636] J. P. Uzan, C. Clarkson, and G. F. R. Ellis, Time drift of cosmological redshifts as a test of the Copernican principle, *Phys. Rev. Lett.* **100** (2008), 191303.
- [637] T. Koivisto and D. F. Mota, Anisotropic dark energy: dynamics of the background and perturbations, *JCAP* **6** (2008), 18.
- [638] D. C. Rodrigues, Anisotropic cosmological constant and the CMB quadrupole anomaly, *Phys. Rev. D* **77** (2008), 023534.
- [639] Gaia website at www.rssd.esa.int/Gaia/
- [640] F. Ding and R. A. C. Croft, Future dark energy constraints from measurements of quasar parallax: Gaia, SIM and beyond, arXiv:astro-ph/0903.3402 (2009).
- [641] B. A. Bassett and M. Kunz, Is cosmic distance-duality violated? *Phys. Rev. D* **69** (2004), 101305.
- [642] A. Avgoustidis, L. Verde, and R. Jimenez, Consistency among distance measurements: transparency, BAO scale and accelerated expansion, *JCAP* **0906** (2009), 012.
- [643] S. Cole, K. B. Fisher, and D. H. Weinberg, Fourier analysis of redshift space distortions and the determination of Ω , *Mon. Not. Roy. Astron. Soc.* **267** (1994), 785.
- [644] P. J. E. Peebles, Tests of cosmological models constrained by inflation, *Astrophys. J.* **284** (1984), 439.
- [645] M. S. Turner, G. Steigman, and L. M. Krauss, Flatness of the universe – Reconciling theoretical prejudices with observational data, *Phys. Rev. Lett.* **52** (1984), 2090.
- [646] N. Sugiyama, N. Gouda, and M. Sasaki, Constraints on universe models with cosmological constant from cosmic microwave background anisotropy, *Astrophys. J.* **365** (1990), 432.
- [647] L. M. Krauss and M. S. Turner, The cosmological constant is back, *Gen. Rel. Grav.* **27** (1995), 1137.
- [648] J. Frieman, M. Turner, and D. Huterer, Dark energy and the accelerating universe, *Ann. Rev. Astron. Astrophys.* **46** (2008), 385.

Index

- action principle, 110
- adiabatic condition, 72, 103, 298
- affine parameter, 76, 289
- age of the Universe, 85, 416
- Alcock–Paczynski effect, 393, 401
- angular diameter distance, 21, 93
 - comoving, 98
- anisotropic models, 424
- anisotropic stress, 69, 300, 333, 409
- anthropic principle, 124
- anti de Sitter (AdS) bulk, 277
- anti de Sitter (AdS) vacuum, 118, 121
- Anti de Sitter/Conformal Field Theory (AdS-CFT)
 - correspondence, 174
- anti-symmetric tensor, 125
- Aristotle, ix
- assisted inflation, 150
- assisted quintessence, 150
- atomic clock, 204
- attractor, 142, 142, 150, 182, 225, 264, 274
- autonomous equations, 180, 191, 219, 271
- average
 - ensemble, 28
 - sample, 28
- axion, 18

- backreaction, 293
- baryon acoustic oscillations (BAO), 5, 102, 282, 393
 - effective distance, 104, 198
 - relative distance, 104
- baryon-dominated epoch, 196
- baryon–photon plasma, 55
- baryonic/dark matter density, 252, 268
- baryons, 1, 63, 69
 - big bang nucleosynthesis, 16
 - coupling to, 190
 - decoupling, 194
 - perturbations, 55
- Bayes’ theorem, 357
- Bayesian approach, 356
 - confidence regions, 359
 - estimators, 362
 - evidence, 358
 - forecasts, 368
 - Jeffrey’s scale, 366
 - marginalization, 360
 - model selection, 363
 - nuisance parameters, 360
 - posterior probability, 358
 - prior probability, 358
- Bessel functions, 96
- bias, 57, 397
 - from non-linearity, 347
 - in coupled models, 312
 - in the growth rate, 398
 - Kaiser’s formula, 83
- Big Bang Nucleosynthesis (BBN), 15, 151
- big-rip, 25, 187, 230
- bispectrum, 347
- black body radiation, 1
- black hole, 416
- Boltzmann equation, 64, 67, 93
- Boltzmann’s constant, 6
- Bose–Einstein distribution, 13, 65
- bosonic string, 174
- bosons, 14, 116
- brane tension, 123
- braneworld, 4, 277, 330
- Brans–Dicke parameter, 258, 260, 267, 326
- Brans–Dicke theory, 5, 206, 256, 258, 333
- bremsstrahlung emission, 411
- broken supersymmetry, 118
- bulk flow, 58

- Calabi–Yau (CY) manifold, 121
- Calan/Tololo Supernova Survey, 90
- cascading gravity, 130
- Cassini tracking, 213
- chameleon mechanism, 195, 205, 306
 - in $f(R)$ gravity, 248
 - in scalar-tensor models, 267
- Chandra satellite, 412, 414
- Chandrasekhar limit, 87
- Chaplygin gas, 225, 308
- chemical potential, 13
- chiral field, 117, 160

- Christoffel symbol, 8
 - perturbed, 44
- COBE, 2, 15, 71, 93
- coincidence problem, 3, 114, 182, 428
- Cold Dark Matter (CDM), 17, 23, 74, 106
 - perturbations, 68, 72, 310
- collision term, 64
- comoving distance, 12, 19, 45
- comoving gauge, 41
- comoving wavenumber, 49, 178, 227, 256
- compactification, 118, 120, 128
 - flux, 121
- complex structure moduli, 121
- Compton scattering, 66, 67
- conditional probability, 358
- conformal anomaly, 232
- conformal factor, 162, 249, 258
- conformal time, 41, 77, 384
- conformal transformation, 162, 173, 248, 258
- conservation equation, *see* continuity equation
- continuity equation, 10, 135, 280, 337
 - perturbed, 48
- Copernican principle, 292
- correlation function, 28, 31, 395
 - in redshift space, 401
- cosmic distance, 18
- Cosmic Lens All-Sky Survey (CLASS), 421
- Cosmic Microwave Background (CMB), 2, 15, 79, 93, 152
 - acoustic peak, 99
 - anisotropies, 5
 - peak position, 95
 - power spectrum, 97, 330, 389
 - shift parameter, 98, 193, 198, 282
 - variance, 97
- cosmic shear, 403
- cosmic time, 7
- cosmic transparency, 426
- cosmological constant, 3, 11, 55, 56, 109, 125
- cosmological constant boundary, 188
- cosmological constant boundary crossing, 266
- cosmological density, 13, 114, 154, 248, 252
- Coulomb scattering, 67
- counts in cells, 37
- coupled dark energy, 189, 196, 215
 - higher-order perturbations, 345
 - parametrization, 196
- coupled quintessence, 163, 201, 205, 222
- coupling strength, 190, 197, 202, 205, 215, 249, 259, 306
- covariant derivative, 10
- cross-correlation spectrum, 387
- crossover scale, 280
- curvature of the Universe, 7, 25, 87, 89, 278
- curvature perturbations, 98
- cycloid equation, 288, 348
- D-brane, 121, 123, 174, 277
 - anti, 120, 174
 - BPS, 174
 - non-BPS, 174
- D-term, 154
- dark energy, 1, 13, 18, 22
 - clustering, 350
 - effective mass, 308
 - N-body simulations, 354
 - papers, 427
 - perturbation, 75, 391
- dark matter, 1, 17, 69
 - candidate, 17
 - fuzzy, 316
 - perturbation, 55
- de Sitter
 - point, 236, 240, 251, 255, 262, 273
 - solution, 112, 158, 173, 281
 - vacuum, 3, 121
- decaying mode, 53
- deceleration parameter, 166
- decoupling epoch, 2, 93
- decoupling of the cosmological constant, 129
- degravitation of the vacuum, 130
- density contrast, 29, 46
 - critical value, 349
 - for a scalar field, 307
- density parameter, 10
- deuterium, 16, 151
- deviation parameter, 238
- dilaton
 - field, 121, 160
 - gravity, 5, 189, 258, 260
- dilatonic ghost condensate model, 173, 179, 217
- dipole moment, 66, 68, 95
- Dirac's delta function, 31, 280, 455
- Dirac–Born–Infeld (DBI)
 - action, 174
 - field, 176
- distortion tensor, 82
- distribution function, 13, 63, 64, 351
- Doppler effect, 12
- Doppler shift, 80
- double exponential potential, 149
- drag epoch, 103
- duality relation, *see* Etherington relation
- Dvali–Gabadadze–Porrati (DGP) model, 5, 277
 - perturbations, 330, 333
- early dark energy, 149
- effective equation of state, 54, 138, 145, 191, 219, 255
 - for $f(R)$ models, 238
 - in Gauss–Bonnet models, 271
 - in Palatini formalism, 262
- effective gravitational constant, 247, 267, 326, 333
- effective potential, 201, 207, 249, 274, 313
- effective string action, 161, 173
- EGRET experiment, 188
- eigenvalues, 140, 141, 191, 219, 240, 255, 263, 274, 372
- eigenvectors, 141
- Einstein, 111, 427
- Einstein equations, 8, 110, 132
 - averaging, 293
 - perturbed, 44

- Einstein frame, 173, 189, 212, 248, 249, 258, 326
- Einstein tensor, 8
 - perturbed, 45
- Einstein–de Sitter universe, 386
 - non-linear perturbations, 340
 - spherical collapse, 348
- electromagnetic field, 203
- electron, 2, 15, 64, 66, 93, 103
- energy density, 9, 14, 136, 175
- energy-momentum tensor, 4, 8, 134, 136, 175
 - perturbed, 47
- entropy, 15, 26
- equation of state, 5, 11, 23, 165
 - k-essence, 175
 - phantom, 266
 - quintessence, 138
- equivalence principle, 60, 160, 211, 251, 267
- ergodic system, 29
- ESSENCE, 92
- Etherington relation, 22, 414, 426
- Euler equation, 49, 337, 345
 - non-linear, 342
- European Extremely Large Telescope (EELT), 422
- exponential potential, 139, 163, 217, 270
- extended supergravity models, 158
- $f(\mathcal{G})$ gravity, 276
- $f(R)$ gravity, 4, 189, 234, 258
- F-term, 156
- Fermi–Dirac distribution, 13, 200
- fermion, 14, 116, 153
- fermion condensate model, 153
- field potential, 117, 122, 135, 153, 174
- fifth force, 205, 210
- figure of merit, 373
- fine structure constant, 203
- fine tuning problem, 113, 114, 428
- fingers-of-god, 59, 400
- Fisher matrix, 356, 368, 396
 - addition of, 375
 - for Gaussian data, 379
 - for weak lensing, 408
 - maximization, marginalization, 371
 - of power spectrum, 377, 391, 395
 - transformation of variables, 370
- five-dimensional
 - bulk, 277
 - Einstein equations, 278
 - Einstein tensor, 279
 - Planck mass, 278
- fixed points, 139, 180, 183, 191, 218, 239, 255, 262, 271
 - instantaneous, 143
 - stability, 140
- flat universe, 85, 281, 282
- flux integers, 124
- 4-form field, 126
- four-velocity, 9, 45, 79
- Fourier space, 31, 49, 65, 325, 342, 406
 - discretization, 377
- Fourier transformation, 31, 331, 341
- free streaming, 96, 316
- frequentist approach, 357
- Friedmann–Lemaître–Robertson–Walker (FLRW)
 - spacetime, 7, 18
- future singularities, 230
- GAIA satellite, 425
- galaxy
 - ages, 418
 - clustering, 106
 - density contrast, 60
 - distribution function, 406
 - merging, 61
 - number counts, 418
 - power spectrum, 107
 - velocity distribution, 400
- galaxy clusters, 411
 - abundance, 414
 - density profile, 411
 - depletion parameter, 414
 - gas fraction, 414
 - hydrostatic equilibrium, 413
 - isothermal distribution, 413
 - mass-temperature relation (M-T), 412
 - sub-structure, 421
- Gamma-ray burst (GRB), 415
- gauge coupling, 160
- gauge field, 203
- gauge transformations, 41
- gauge-invariant variables, 41, 300
- Gauss’s theorem, 110, 249
- Gauss–Bonnet dark energy, 269
- Gauss–Bonnet (GB) term, 131, 269
- Gaussian initial conditions, 47, 339
- General Relativity, 3, 111, 235
- generalized Chaplygin gas (GCG) model, 225
- geodesic equation, 19, 64, 76, 289
- ghost, 25, 162, 177, 186, 269, 276, 282
- ghost condensate model, 173
- global supersymmetric theory, 153
- globular clusters, 85, 417
- gluino condensation, 118
- Gold data, 92, 275
- gravitational constant, 6
 - bare, 234
 - variation, 195
- gravitational instability, 53, 227
- gravitational lensing
 - giant arcs, 421
 - mass estimations, 412
 - strong, 418
 - time delay, 418
- gravitational potential, 69, 72
 - large-scale solution, 51
 - local value, 247
 - present value, 74
 - small-scale solution, 56
- gravitational waves, 43
- gravitino, 157
- gravity mediation, 160
- growing mode, 51, 53

- growth function, 47, 74, 387, 390, 394
 - higher-order, 339
- growth index, 56
- growth rate, 56, 303, 304, 395, 398
 - fit, 304
 - from clusters, 412, 414
 - in coupled models, 311
 - in modified gravity, 328
- halo, 355, 419
 - clusters, 419
 - profiles, 419
- helium, 16, 151
- Helmholtz's theorem, 42
- hidden sector, 160
- hierarchy problem, 113
- High-redshift Supernova Search Team (HSST), 1, 87
- higher-order moments, 31
- hot Big Bang model, 7, 112
- Hot Dark Matter (HDM), 17
- Hubble constant, 12, 84, 101, 113, 158
- Hubble Key Project, 13, 101
- Hubble parameter, 9, 88, 416
 - for big-rip models, 25
 - in braneworld models, 280
 - in the redshift drift, 423
 - measures the age of the Universe, 84
 - parametrization, 198
 - radial, 286
 - reconstruction of, 163
 - transverse, 286
- Hubble radius, 51, 68, 179, 281
 - present, 13
- Hubble radius crossing, 71
- Hubble Space Telescope (HST), 92
- Hubble's law, 12, 20, 112
- inflation, 1, 11, 71, 125, 236
- infra-red (IR) instability, 179
- initial condition, 70
 - adiabatic, 80, 306
- instant preheating, 151
- instantaneous fixed point, 263
- instantaneous minima, 265, 327
- Integrated Sachs–Wolfe (ISW) effect, 74, 79, 80, 93, 384, 389
 - cross-correlation, 306
 - for $f(R)$ gravity, 323
 - for early acceleration, 196
- intra-cluster medium, 411
- intrinsic temperature fluctuation, 80
- inverse power-law potential, 144, 154, 168, 204, 213
- Jacobian matrix, 191, 192, 219, 263, 274
 - to transform the Fisher matrix, 370
- Jeans length, 52, 227
- Jordan frame, 206, 212, 259, 326, 327
- K-correction, 88
- k-essence, 4, 172, 182, 228
 - attractor, 184
- Kachru–Kallosh–Linde–Trivedi (KKLT) model, 121
- Kähler modulus, 121
- Kähler potential, 117, 155
- Kaluza–Klein theories, 277
- kinetic energy
 - as a fixed point, 140
 - for a chameleon field, 209
 - for phantom fields, 186
 - in k-essence models, 172
 - in spherical collapse, 349
 - in unified models, 229
 - of a scalar field, 4, 215
- kinetic points, 221, 262
- Klein–Gordon equation, 136
- Lagrangian density, 109, 173, 190, 215, 228
- Λ CDM model, 4, 74, 106, 109, 167, 200, 241
- large extra dimensions, 277
- Large Hadron Collider (LHC), 2
- large-scale structure, 70, 106
- last scattering surface, 2
- Legendre polynomial, 66, 455
- Lemaître's model, 112
- Lemaître–Tolman–Bondi (LTB) metric, 286
- lensing potential, 81, 335, 409
- Levi-Civita tensor, 43
- LHC experiments, 18
- likelihood function, 357, 369
- likelihood method, 356, 359, 367
- Limber equation, 405
- linear expansion, 247, 251
- local gravity constraints, 214, 236, 245, 252, 267, 276
- local gravity experiments, 195, 205
- long-range force, 4, 153, 189
- longitudinal component, 42
- longitudinal gauge, 41, 44
- loop quantum cosmology, 233
- low-energy effective string theory, 173
- luminosity distance, 5, 20, 87, 163, 198
- Lyman- α lines, 423
- Mach's principle, 257
- Markov chain, 367
- mass function, 352, 410
 - formula, 353
- mass varying neutrino, 200
- massive neutrino, 56, 194, 202, 398
- massless
 - field, 203, 267
 - limit, 247, 325, 326, 333
 - neutrino, 67, 69
 - particles, 63
- matter Lagrangian, 129, 205, 258
- matter perturbations, 55
 - at second order, 338
 - growth rate, 395
 - in $f(R)$ gravity, 317
 - in scalar-tensor models, 324
 - in the braneworld model, 330
 - in the Palatini formalism, 257
- matter point, 142, 180, 240, 255, 266

- matter power spectrum, 70, 323, 394, 395
 - biased, 107
 - in $f(R)$ gravity, 328
 - in the Chaplygin model, 227
 - peak wavelength, 75
- matter-dominated epoch, 2, 71, 73, 192, 236
- membrane, 126
- metric, 7
- metric formalism, 234, 245, 264
- Minkowski background, 177, 206, 267
- Minkowski bulk, 277
- modified matter models, 134, 172
- modulus field, 118, 119
- moments
 - second-order, 39
 - third-order, 39
- momentum vector, 76
- monopole moment, 66, 94
- multipole moments, 68, 97

- N-body simulations, 61, 353, 396, 409
- Navarro–Frenk–White profile, 421
- negative instability, 179
- neutralino, 18
- neutrinos, 2, 15, 63, 69, 70
- neutrons, 16, 151
- Neveu–Schwarz (NS)–NS flux, 121
- Newtonian gauge, 41, 42, 44
- Newtonian gravity, 11, 336
- Newtonian limit, 52
- Newtonian regime, 57
- no-go theorem, 120
- no-scale models, 119, 158
- non-Gaussianity, 345, 408
- non-linear regime, 31, 49, 248, 251
- non-relativistic matter, 13, 16, 49, 67, 135, 189
- nonminimally coupled scalar field, 260
- normal ordering, 129
- number of e-foldings, 53, 138

- Oklo natural fission reactor, 203
- open universe, 86, 90
- optical depth, 66, 96, 384

- Palatini formalism, 235, 253
- parametrization, 93, 164, 282, 370
 - logarithmic, 165
 - redshift, 165
 - scale factor, 165
- particle physics, 3, 113, 144, 153
- Peccei–Quinn (PQ) symmetry, 158
- peculiar velocity, 12, 45, 56, 57
 - redshift distortion, 58
- perfect fluid, 46, 135, 302, 337
- perfect fluid model, 4, 225
- perturbations, 41
 - degrees of freedom, 305
 - entropy, 305
 - equations in General Relativity, 50
 - general fluid, 302
 - in $f(R)$ gravity, 317, 324
 - in DGP model, 330
 - large-scale limit, 300
 - linear, 296
 - metric, 297, 333
 - perfect fluid equations, 299
 - scalar field, 306
 - scalar-tensor gravity, 324
 - second-order, 336
 - small scales, 303
 - small-scale limit, 310
 - spherical, 338
 - stability, 318
- perturbed Hamiltonian, 178
- perturbed metric, 41, 42, 43, 63
- phantom, 25, 105, 173, 178, 186
- ϕ MDE, 192, 221, 239, 262
- photometric redshift, 395
- photon propagation, 76, 335, 384
- photon–baryon plasma, 70
- photons, 2, 14, 16, 63, 76
- physical distance, 12
- pivot point, 371
- Planck mass, 6
- Planck scale, 113, 159
- Planck’s constant, 6
- Plato, ix
- Poisson equation, 50, 310, 336, 386, 406
 - modified, 345
- Poissonian, 33
- polarization, 76
- positron, 15
- post-Newtonian parameter, 212, 247, 267
- power spectrum, 31, 71, 330, 388, 394, 405
 - higher-order, 346
 - noise, 36, 404
 - non-linear, 354
 - normalization, 39
- Press–Schechter theory, 349, 351, 421
- pressure, 9, 14, 136, 175
- principal component analysis (PCA), 380
- probability distribution function (PDF), 356, 369
- protons, 16, 64, 151
- Pseudo-Nambu–Goldstone Boson (PNGB), 138, 158

- quadrupole, 70, 72
- quantum
 - chromodynamics (QCD), 18, 113, 153
 - corrections, 161, 232
 - cosmology, 125
 - field theory, 116, 129
 - fluctuations, 71, 178
 - operators, 116
 - tunneling effect, 126
- quarks, 153
- quasars, 203, 423
- quintessence, 4, 134, 163
 - coupled, 306
 - freezing models, 137, 168
 - mass, 153
 - potential, 164
 - thawing models, 137, 169

- quintessential inflation, 151
- quintom, 188
- radiation point, 180, 243, 255
- radiation-dominated epoch, 2, 55, 72, 192
- radiation-matter equality, 23, 72, 73, 99
- radiative corrections, 159
- Ramond–Ramond (R-R) flux, 121
- real-time cosmology, 422
- reciprocity relation, *see* Etherington relation
- recombination, 67, 93
- reconstruction, 163
- redshift, 12
 - distortion, 59, 61, 394, 398
 - drift, 292, 422
 - surveys, 398
- reduced Planck mass, 6
- Regge slope parameter, 161
- reheating, 151
- relativistic particle, 10, 13
- Ricci scalar, 8, 110, 234
 - perturbed, 45
- Ricci tensor, 8
 - for LTB metric, 287
 - perturbed, 45
- rotational modes, 43
- runaway dilaton scenario, 161
- Sachs–Wolfe (SW) effect, 78, 79, 80
- saddle point, 141, 192, 221, 240, 255, 274
- Sandage effect, *see* redshift drift
- scalar curvature, 8
- scalar field, 4, 13, 135, 172, 189, 215, 257, 269
 - as dark matter, 309
 - Compton wavelength, 316
 - energy-momentum tensor, 307
 - oscillations, 316
- scalar-field dominated solution, 140, 219, 262
- scalar-tensor theories, 4, 189, 258
 - perturbations, 324
- scalatron, 235
- scale factor, 7, 12
- scale-invariant spectrum, 71, 98
- scaling attractor, 193, 195, 223
- scaling Lagrangian, 217
- scaling matter era, 181, 272, 274
- scaling solution, 140, 149, 177, 215, 220, 262, 272
- Schwinger pair creation, 126
- shear-free gauge, 44
- shot-noise, 394
- single fluid, 50, 225
- singular isothermal sphere (SIS), 419
- singularity, 223, 253
 - type I, 230
 - type II, 230
 - type III, 231
 - type IV, 231
- skewness, 343, 344, 346
- SLACS, 421
- Sloan Digital Sky Survey (SDSS), 5, 38, 102, 104, 388, 392
- slow-roll parameters, 137
- smoothing, 164, 344, 351
- solar mass, 420
- solar system constraints, 212, 213
- sound horizon, 94, 98, 193
- sound speed, 47, 52, 68, 69, 94, 178, 218, 227, 230
 - adiabatic, 47, 298
 - in spherical collapse, 351
 - non-adiabatic, 47
 - of a scalar field, 307, 314
 - of dark energy, 390
 - total, 298, 301
- speed of light, 6
- spherical collapse, 347, 351
- spherical harmonics, 96
- spherically symmetric body, 207, 246
- stability conditions for k-essence, 177
- stability of tensor perturbations, 276
- stable node, 141
- stable spiral, 141
- standard candle, 87, 413
 - GRBs, 416
- statefinder, 166
- stationary solution, 312
- statistical homogeneity, 33
- string coupling, 161
- string landscape, 4, 126, 128
- string mass, 160
- string theory, 116, 128, 157, 160, 277
- strong anthropic principle, 124
- strong CP problem, 158
- strong gravitational background, 253
- structure formation, 2, 17
- sub-horizon
 - approximation, 75, 256
 - scales, 51, 54
- sudden future singularity, 230
- Sunyaev–Zel’dovich effect, 290, 294
- super-horizon
 - approximation, 72
 - scales, 51, 69, 71, 299
- supergravity, 116, 154
 - effective action, 118
- superluminal propagation, 179, 185
- supernova, 87
 - absolute magnitude, 88
 - apparent magnitude, 87
 - flux, 87
 - type Ia (SN Ia), 1, 87, 244
 - type Ib, 87
 - type II, 87
- Supernova Cosmology Project (SCP), 1, 87
- SuperNova Legacy Survey (SNLS), 92, 198
- superpotential, 117, 154
- supersymmetric particles, 18
- supersymmetric theories, 18, 116
- supersymmetry, 113, 116, 156
- supersymmetry breaking, 155
- Swiss-cheese model, 291

- tachyon field, 174, 175, 217
- temperature perturbation, 65
- tensor perturbation, 276
- thick-shell, 211
- thin-shell, 205
- thin-shell parameter, 211, 251
- Thomson cross section, 66
- Thomson scattering, 2
- three-cycles, 128
- tight coupling, 103
 - limit, 68, 95
- top-hat filtering, 36
- total equation of state, *see* effective equation of state
- total matter perturbation, 50, 60, 227
- trace, 189, 235
- traceless, 189
- tracker solution, 145
- tracking condition, 143
- transfer function, 73, 74
 - Bardeen–Bond–Kaiser–Szalay (BBKS), 74
- transverse component, 42
- two-degree Field (2dF) Galaxy Redshift Survey, 38, 104, 228, 392
- ultra-violet (UV) instability, 177, 220
- unbroken supersymmetry, 157
- unified models of dark energy and dark matter, 225, 228
- unstable node, 141
- unstable spiral, 141
- vacuum energy, 3, 113, 117, 125
- vacuum expectation value, 160
- vacuum instability, 187
- varying α , 203
- velocity divergence, 46
- velocity field, 56, 60, 62
- virialization, 349
 - in clusters, 412
 - radius, 349
- viscosity, 338
- visibility function, 96
- void model, 286
- warp factor, 122, 174
- warped compactification, 122
- weak anthropic principle, 124
- weak coupling limit, 162
- weak energy condition, 164
- weak lensing, 78, 81, 335, 403, 409
 - convergence, 82
 - convergence spectrum, 407
 - distortion tensor, 404
 - galaxy ellipticity, 403
 - power spectrum, 408
 - shear field, 82
 - transformation matrix, 81
- Weakly Interacting Massive Particles (WIMPs), 18
- Weinberg’s bound, 125
- Weyl tensor, 280
- white dwarf, 85
- Wiener–Khinchin theorem, 32
- Wilkinson Microwave Anisotropy Probe (WMAP), 2, 17, 71, 86, 93, 99, 384
- window function, 34, 58, 345
 - top-hat function, 34
- X-ray spectra, 412
- XMM-Newton satellite, 412
- Yang–Mills coupling, 174
- Yukawa term, 314, 327
- zero-point energy, 3, 113