Lecture notes: Advanced Cosmology

Luca Amendola
of Heidelberg
.amendola@thphys.uni-heidelberg.de
://www.thphys.uni-heidelberg.de/~amendola/teaching.html

Caroline Heneka
of Heidelberg
c.heneka@thphys.uni-heidelberg.de

v. 2.1

July 25, 2023

Contents

[a] Text adapted from Amendola et al. *Universe*, vol. 6, p. 20

[b] Adapted from Amendola & Tsujikawa, *Dark Energy. Theory and Observations*, CUP 2010.

Acknowledgments and credits

This course is addressed to master students; we will make often use of concepts from General Relativity and some basic astronomy. The student will benefit a lot by reading the relevant chapters in the following texts and in astrophysics textbooks.

Suggested readings:
L. Amendola, *Lecture Notes on Cosmology* (all the needed background material is here)
S. Dodelson, *Modern Cosmology*, Academic Press (my favourite)
L. Amendola & S. Tsujikawa, *Dark Energy. Theory and Observations*, CUP (more advanced material; some text in these Lecture Notes is adapted directly from this book)
O. Piattella, *Lecture Notes in Cosmology*, Springer (recent and complete)
M. Bartelmann, *Observing the Big Bang*, Lecture notes
M. Bartelmann, *Cosmology*, Lecture notes
B. M Schaefer, *Cosmology*, Lecture notes
More specialized texts:
N. Sugiyama, *Introduction to temperature anisotropies of CMB*, PTEP,2014, 06B101
D. Weinberg et al., *Observational probes of cosmic acceleration*, arXiv:1201.2434

Part I

Theory

Chapter 1

Review of standard cosmology

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. No new concept with respect to the standard Cosmology course is introduced.

1.1 Einstein-Hilbert Lagrangian

Einstein's equations *in vacuo* can also be obtained by varying a gravitational action, called Hilbert-Einstein action

$$A = \int \sqrt{-g} R d^4x \tag{1.1}$$

In fact, we note that the metric determinant transforms as

$$g' = gJ^{-2}$$

where $J_{\mu\nu} \equiv \partial y_\mu / \partial x_\nu$ is the Jacobian of the general transformation that brings us from g to g'. It is clear then that $\sqrt{-g'}d^4x = \sqrt{-gJ^{-2}}|J|d^4y = \sqrt{-g}d^4y$ is invariant under general transformations: this explains the factor $\sqrt{-g}$ in the action. By varying A with respect to the metric and using the relation

$$\frac{\partial R}{\partial g_{\mu\nu}} = \frac{\partial (g_{\alpha\beta}R^{\alpha\beta})}{\partial g_{\mu\nu}} = R^{\mu\nu} + g_{\alpha\beta}\frac{\partial R^{\alpha\beta}}{\partial g_{\mu\nu}}$$

and also

$$\delta\sqrt{-g} = -\frac{1}{2}\sqrt{-g}(\delta g^{\mu\nu})g_{\mu\nu}$$

we obtain

$$\delta A = \int \sqrt{-g} d^4x [-\frac{1}{2}g_{\mu\nu}R + R_{\mu\nu} + g^{\alpha\beta}\frac{\partial R_{\alpha\beta}}{\partial g_{\mu\nu}}]\delta g^{\mu\nu} = 0 \tag{1.2}$$

We can now show that the term

$$\delta A = \int \sqrt{-g} d^4x [g^{\alpha\beta}\frac{\partial R_{\alpha\beta}}{\partial g_{\mu\nu}}]\delta g^{\mu\nu} = 0 \tag{1.3}$$

is a total differential (i.e. $\delta A = 0$ is an identity) and is therefore irrelevant for as concerns the equation of motion. In fact one can write

$$R_{\mu\nu} = \Gamma^{\alpha}_{\mu\nu;\alpha} - \Gamma^{\beta}_{\beta\mu;\nu} \tag{1.4}$$

(where the covariant derivative is to be meant only wrt the upper index of the Christoffel symbols) and

$$\sqrt{-g}g^{\mu\nu}\delta R_{\mu\nu} = \sqrt{-g}(g^{\mu\nu}\delta\Gamma^{\alpha}_{\mu\nu} - g^{\mu\alpha}\delta\Gamma^{\beta}_{\beta\mu})_{;\alpha} \tag{1.5}$$

The term inside parentheses is the covariant derivative of the vector $V^{\alpha} \equiv g^{\mu\nu}\delta\Gamma^{\alpha}_{\mu\nu} - g^{\mu\alpha}\Gamma^{\beta}_{\beta\mu}$ and can therefore be written as

$$(V^{\alpha}\sqrt{-g})_{,\alpha} \tag{1.6}$$

(notice now the derivative is the ordinary one))i.e. as a total derivative.

Then the Einstein equations in vacuum follow

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 0$$

1.2 Friedmann equations

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

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^{\mu}\mathrm{d}x^{\nu} = -\mathrm{d}t^2 + a^2(t)\mathrm{d}\sigma^2\,, \tag{1.7}$$

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

$$\mathrm{d}\sigma^2 = \gamma_{ij}\mathrm{d}x^i\mathrm{d}x^j = \frac{\mathrm{d}r^2}{1-Kr^2} + r^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2)\,. \tag{1.8}$$

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. (1.7) the Greek indices μ and ν run from 0 to 3, whereas in Eq. (1.8) 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 (1.7) from a maximally symmetric spacetime. In addition to the cosmic time t, we also introduce the conformal time η defined by

$$\eta \equiv \int a^{-1}\,\mathrm{d}t\,. \tag{1.9}$$

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:

$$\Gamma^{\mu}_{\nu\lambda} = \frac{1}{2}g^{\mu\alpha}(g_{\alpha\nu,\lambda} + g_{\alpha\lambda,\nu} - g_{\nu\lambda,\alpha})\,, \tag{1.10}$$

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

$$R_{\mu\nu} = \Gamma^{\alpha}_{\mu\nu,\alpha} - \Gamma^{\alpha}_{\mu\alpha,\nu} + \Gamma^{\alpha}_{\mu\nu}\Gamma^{\beta}_{\alpha\beta} - \Gamma^{\alpha}_{\mu\beta}\Gamma^{\beta}_{\alpha\nu}\,. \tag{1.11}$$

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

$$R = g^{\mu\nu}R_{\mu\nu}\,. \tag{1.12}$$

We can then evaluate the Einstein tensor

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R \,. \tag{1.13}$$

The cosmological dynamics can be obtained by solving the Einstein equations

$$G^{\mu}_{\nu} = 8\pi G T^{\mu}_{\nu} \,, \tag{1.14}$$

where T^{μ}_{ν} is the energy-momentum tensor of matter components. The l.h.s. of Eq. (1.14) 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 (1.7) the non-vanishing components of Christoffel symbols are

$$\Gamma^0_{ij} = a^2 H \gamma_{ij} \,, \quad \Gamma^i_{0j} = \Gamma^i_{j0} = H \delta^i_j \,, \tag{1.15}$$

$$\Gamma^1_{11} = \frac{Kr}{1-Kr^2} \,, \quad \Gamma^1_{22} = -r(1-Kr^2) \,, \quad \Gamma^1_{33} = -r(1-Kr^2)\sin^2\theta \,, \tag{1.16}$$

$$\Gamma^2_{33} = -\sin\theta\cos\theta \,, \quad \Gamma^2_{12} = \Gamma^2_{21} = \Gamma^3_{13} = \Gamma^3_{31} = \frac{1}{r} \,, \quad \Gamma^3_{23} = \Gamma^3_{32} = \cot\theta \,, \tag{1.17}$$

where

$$H \equiv \dot{a}/a \,. \tag{1.18}$$

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. (1.16) and (1.17) correspond to those for the three-dimensional metric (1.8) with the curvature K.

From Eqs. (1.11) and (1.12) the Ricci tensor and the scalar curvature are

$$R_{00} = -3\left(H^2 + \dot{H}\right) \,, \quad R_{0i} = R_{i0} = 0 \,, \quad R_{ij} = a^2\left(3H^2 + \dot{H} + 2K/a^2\right)\gamma_{ij} \,, \tag{1.19}$$

$$R = 6\left(2H^2 + \dot{H} + K/a^2\right) \,. \tag{1.20}$$

From Eq. (1.13) together with the relation $G^{\mu}_{\nu} = g^{\mu\alpha} G_{\alpha\nu}$, the Einstein tensor is

$$G^0_0 = -3\left(H^2 + K/a^2\right) \,, \quad G^0_i = G^i_0 = 0 \,, \quad G^i_j = -\left(3H^2 + 2\dot{H} + K/a^2\right)\delta^i_j \,. \tag{1.21}$$

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

$$T^{\mu}_{\nu} = (\rho + P) u^{\mu} u_{\nu} + P \delta^{\mu}_{\nu} \,, \tag{1.22}$$

where $u^{\mu} = (-1, 0, 0, 0)$ is the four-velocity of the fluid in comoving coordinates, and ρ and P are function of t. The (00) and (ij) components of T^{μ}_{ν} are $T^0_0 = -\rho$ and $T^i_j = P\delta^i_j$. 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 (1.14) we obtain

$$H^2 = \frac{8\pi G}{3}\rho - \frac{K}{a^2} \,, \tag{1.23}$$

$$3H^2 + 2\dot{H} = -8\pi G P - \frac{K}{a^2} \,. \tag{1.24}$$

Eliminating the term K/a^2 gives

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P) \,. \tag{1.25}$$

Multiplying Eq. (1.23) by a^2, differentiating and using Eq. (1.25), we find

$$\dot{\rho} + 3H(\rho + P) = 0 \,. \tag{1.26}$$

The Einstein tensor satisfies the Bianchi identities

$$\nabla_\mu G^\mu_\nu \equiv \frac{\partial G^\mu_\nu}{\partial x^\mu} + \Gamma^\mu_{\alpha\mu} G^\alpha_\nu - \Gamma^\alpha_{\nu\mu} G^\mu_\alpha = 0\,, \tag{1.27}$$

where ∇_μ denotes the covariant derivative. Sometimes we use also the symbol "$_{;\mu}$" to represent the covariant derivative. From the Einstein equations (1.14) it follows that $\nabla_\mu T^\mu_\nu = 0$, which gives the same equation as (1.26) in the FLRW background (see the problem [2.1]). Hence Eq. (1.26) is called conservation or continuity equation.

Equation (1.23) can be written in the form:

$$\Omega_M + \Omega_K = 1\,, \tag{1.28}$$

where

$$\Omega_M \equiv \frac{8\pi G\rho}{3H^2}\,, \qquad \Omega_K \equiv -\frac{K}{(aH)^2}\,. \tag{1.29}$$

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

$$\Omega_r^{(0)} = \frac{8\pi G\rho_r^{(0)}}{3H_0^2}\,, \quad \Omega_m^{(0)} = \frac{8\pi G\rho_m^{(0)}}{3H_0^2}\,, \quad \Omega_{\rm DE}^{(0)} = \frac{8\pi G\rho_{\rm DE}^{(0)}}{3H_0^2}\,, \quad \Omega_K^{(0)} = -\frac{K}{(a_0H_0)^2}\,. \tag{1.30}$$

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[a]. 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 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

$$w \equiv P/\rho\,. \tag{1.31}$$

If w is a constant, one can analytically find the evolution of ρ and a for the flat universe ($K = 0$). Solving Eqs. (1.23) and (1.26) in this case, we obtain the following solutions

$$\rho \propto a^{-3(1+w)}\,, \quad a \propto (t - t_i)^{2/(3(1+w))}\,, \tag{1.32}$$

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. (1.25), i.e.

$$P < -\rho/3 \quad \to \quad w < -1/3\,, \tag{1.33}$$

where ρ is assumed to be positive. This implies that the power $2/(3(1 + w))$ in the scale factor $a(t)$ is larger than 1. 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

[a] See Sec. 1.4 for the definition of cold dark matter.

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. (1.26) 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.

1.3 Hubble's law

In 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

$$z \equiv \frac{\lambda_0}{\lambda} - 1 = \frac{a_0}{a} - 1\,, \tag{1.34}$$

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

$$z \simeq v/c\,. \tag{1.35}$$

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

$$\dot{\boldsymbol{r}} = H\boldsymbol{r} + a\dot{\boldsymbol{x}}\,. \tag{1.36}$$

The velocity $\boldsymbol{v}_H \equiv H\boldsymbol{r}$ appears because of the presence of the cosmic expansion. On the other hand, the velocity $\boldsymbol{v}_p \equiv a\dot{\boldsymbol{x}}$, 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

$$v \equiv \dot{\boldsymbol{r}} \cdot \boldsymbol{r}/r = Hr + \boldsymbol{v}_p \cdot \boldsymbol{r}/r\,, \tag{1.37}$$

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

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

$$v \simeq H_0 r\,. \tag{1.38}$$

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 (1.38) 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

$$H_0 = 100\,h\ \mathrm{km\,sec^{-1}\,Mpc^{-1}} = 2.1332h \times 10^{-42}\ \mathrm{GeV}\,, \tag{1.39}$$

where

$$1\ \mathrm{Mpc} = 3.08568 \times 10^{24}\ \mathrm{cm} = 3.26156 \times 10^6\ \text{light years}\,. \tag{1.40}$$

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

$$h = 0.72 \pm 0.08\,. \tag{1.41}$$

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

$$t_H \equiv 1/H_0 = 9.78 \times 10^9\,h^{-1}\ \text{years}\,, \tag{1.42}$$

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

$$D_H \equiv \frac{c}{H_0} = 2998\, h^{-1}\ \mathrm{Mpc}\,, \tag{1.43}$$

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

It is also convenient to introduce the critical density

$$\rho_c^{(0)} \equiv \frac{3H_0^2}{8\pi G} = 1.88\, h^2 \times 10^{-29}\, \mathrm{g\,cm^{-3}}\,, \tag{1.44}$$

which represents the averaged cosmological density in the universe today. Note that we have used Eqs. (1.39) and (1.40) together with the value $G = 6.67 \times 10^{-8}\,\mathrm{cm^3\,g^{-1}\,sec^{-2}}$ to obtain the numerical value in Eq. (1.44). The critical density (1.44) is very small compared to densities in the local structure of the universe ($\rho \simeq 5\,\mathrm{g/cm^3}$ for Earth and $\rho \simeq 10^{-24}\,\mathrm{g/cm^3}$ for the homogeneous baryon/dark matter density in our galaxy). An even smaller fraction is responsible for the present accelerated expansion of the universe.

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

$$f(\boldsymbol{p}) = \frac{1}{\exp(E-\mu)/T \pm 1}\,, \tag{1.45}$$

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 $\boldsymbol{x}$ of the species, but Eq. (1.45) only depends on $p \equiv |\boldsymbol{p}|$ because of the homogeneity of the universe. Since the minimum volume of phase space in terms of $\boldsymbol{x}$ and $\boldsymbol{p}$ is given by $(2\pi\hbar)^3$ due to Heisenberg's principle, the number of phase space elements is $\mathrm{d}^3x\,\mathrm{d}^3p/(2\pi\hbar)^3$. Then the energy density ρ and the pressure P with g_* internal degrees of freedom are [73, 74]

$$\rho = g_* \int \frac{\mathrm{d}^3p}{(2\pi\hbar)^3} E(p)\, f(p) = \frac{g_*}{2\pi^2} \int_m^\infty \mathrm{d}E\, \frac{(E^2-m^2)^{1/2}}{\exp[(E-\mu)/T] \pm 1} E^2\,, \tag{1.46}$$

$$P = g_* \int \frac{\mathrm{d}^3p}{(2\pi\hbar)^3} \frac{pv}{3} f(p) = g_* \int \frac{\mathrm{d}^3p}{(2\pi\hbar)^3} \frac{p^2}{3E}\, f(p) = \frac{g_*}{6\pi^2} \int_m^\infty \mathrm{d}E\, \frac{(E^2-m^2)^{3/2}}{\exp[(E-\mu)/T] \pm 1}\,. \tag{1.47}$$

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. (1.47) 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. (1.47) 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. (1.46) and (1.47) 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 $m \to 0$ in Eqs. (1.46) and (1.47). For non-degenerate particles ($T \gg \mu$) we obtain

$$\rho = \begin{cases} (\pi^2/30) g_* T^4\,, & \text{(Bosons)} \\ (7/8)(\pi^2/30) g_* T^4\,, & \text{(Fermions)} \end{cases} \tag{1.48}$$

$$P = \rho/3\,, \tag{1.49}$$

where we have used $\int_0^\infty {\rm d}^3x\, x^3/(e^x-1) = \pi^4/15$ and $\int_0^\infty {\rm d}^3x\, x^3/(e^x+1) = 7\pi^4/120$. The result (1.49) 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

$$\rho_\gamma = \frac{\pi^2}{15} T^4 \,. \tag{1.50}$$

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

$$\Omega_\gamma^{(0)} \equiv \frac{8\pi G \rho_\gamma^{(0)}}{3H_0^2} = \frac{\rho_\gamma^{(0)}}{\rho_c^{(0)}} = 2.469 \times 10^{-5}\, h^{-2} \,. \tag{1.51}$$

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. (1.32) with $w = 1/3$), the comparison with Eq. (1.50) 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. (1.48) in the fermionic case, the energy density of neutrinos, including anti-particles, is given by

$$\rho_\nu = N_{\rm eff} \frac{7\pi^2}{120} T_\nu^4 \,, \tag{1.52}$$

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

The Big Bang Nucleosynthesis (BBN) occurred around the energy scale $\sim$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 $N_{\rm eff}$. The current standard value is $N_{\rm eff} = 3.04$ [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. (1.50) and (1.52), the relation between the neutrino density and the photon density is $\rho_\nu = N_{\rm eff}(7/8)(4/11)^{4/3}\rho_\gamma$. Hence the present density parameter of radiation, which is the sum of photons and relativistic neutrinos, yields

$$\Omega_r^{(0)} = \frac{\rho_\gamma^{(0)} + \rho_\nu^{(0)}}{\rho_c^{(0)}} = \Omega_\gamma^{(0)}(1 + 0.2271 N_{\rm eff}) \,, \tag{1.53}$$

where $\Omega_\gamma^{(0)}$ is given in Eq. (1.51). If we take the values $h = 0.72$ and $N_{\rm eff} = 3.04$, 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. (1.46) and (1.47) reduce to

$$\rho = g_*\, m \left(\frac{mT}{2\pi} \right)^{3/2} \exp\left[-(m-\mu)/T \right] \,, \tag{1.54}$$

$$P = g_*\, T \left(\frac{mT}{2\pi} \right)^{3/2} \exp\left[-(m-\mu)/T \right] = \frac{T}{m}\rho \,, \tag{1.55}$$

which are valid for both bosonic and fermionic particles. See the problem [2.3] for the derivation of (1.54) and (1.55). This means 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:

$$\Omega_b^{(0)} h^2 = 0.020 \pm 0.002 \qquad \text{(BBN constraint)}, \tag{1.56}$$

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. Crudely speaking the perturbations in CMB anisotropies with comoving wavenumber k satisfy the equation for the harmonic oscillator, $\frac{\mathrm{d}^2 x}{\mathrm{d}\eta^2} + k^2 c_s^2 x = 0$, with some corrections. 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 the present density parameter of baryons is [15]

$$\Omega_b^{(0)} h^2 = 0.02267^{+0.00058}_{-0.00059} \qquad \text{(WMAP 5 year constraint)}\,, \tag{1.57}$$

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. (1.57). 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 be only 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]

$$\Omega_c^{(0)} h^2 = 0.1131 \pm 0.00034 \qquad \text{(WMAP 5 year constraint)}\,, \tag{1.58}$$

at the 68 % confidence level. For the value $h = 0.72$ we have $\Omega_c^{(0)} = 0.2182$ for the central value in Eq. (1.58).

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 the 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.

The 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 examples, 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. (1.53), (1.57), and (1.58) 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:

$$\Omega_{\rm DE}^{(0)} = 0.726 \pm 0.015 \qquad (\text{WMAP 5 year constraint})\,. \tag{1.59}$$

1.5 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 (1.7). 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. (1.8), the 3-dimensional space line-element is expressed as

$$d\sigma^2 = d\chi^2 + (f_K(\chi))^2 \left(d\theta^2 + \sin^2\theta\, d\phi^2\right)\,, \tag{1.60}$$

where

$$f_K(\chi) = \begin{cases} \sin\chi & (K = +1)\,, \\ \chi & (K = 0)\,, \\ \sinh\chi & (K = -1)\,. \end{cases} \tag{1.61}$$

The function (1.61) can be written in a unified way:

$$f_K(\chi) = \frac{1}{\sqrt{-K}} \sinh\left(\sqrt{-K}\chi\right)\,, \tag{1.62}$$

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

1.5.1 Comoving distance

Let us first compute the comoving distance d_c. The light travelling along the χ direction satisfies the geodesic equation: $ds^2 = -c^2dt^2 + a^2(t)d\chi^2 = 0$, 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

$$d_c \equiv \chi_1 = \int_0^{\chi_1} d\chi = -\int_{t_0}^{t_1} \frac{c}{a(t)} dt\,. \tag{1.63}$$

From Eq. (1.34) it follows that $dt = -dz/[H(1+z)]$. Then the comoving distance is given by

$$d_c = \frac{c}{a_0 H_0} \int_0^z \frac{d\tilde{z}}{E(\tilde{z})}\,, \tag{1.64}$$

where

$$E(z) \equiv H(z)/H_0\,. \tag{1.65}$$

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

$$\int_0^z \frac{\mathrm{d}\tilde{z}}{E(\tilde{z})} = z - \frac{E'(0)}{2}z^2 + \frac{1}{6}\left\{2E'(0)^2 - E''(0)\right\}z^3 + \mathcal{O}(z^4)\,, \tag{1.66}$$

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

$$d_c \simeq \frac{c}{a_0 H_0}z\,, \quad \text{for} \quad z \ll 1\,. \tag{1.67}$$

On using the relation (1.35), we find

$$v \simeq (a_0 H_0)d_c\,. \tag{1.68}$$

This shows that the recessional velocity v of the object is proportional to d_c with the proportionality constant a_0H_0. For the physical distance $r = ad_c$ we find $r \simeq (c/H_0)z/(1+z) \simeq v/H_0$, which means that Hubble's law (1.38) is satisfied. Hubble's law written as in Eq. (1.38) is valid therefore only in the low-redshift region $z \ll 1$. For $z \gtrsim 1$ the higher-order terms in Eq. (1.66) become important so that Hubble's law is subject to be modified.

1.5.2 Luminosity distance

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

$$d_L^2 \equiv \frac{L_s}{4\pi\mathcal{F}}\,, \tag{1.69}$$

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 (1.69) yields

$$d_L^2 = (a_0 f_K(\chi))^2\,\frac{L_s}{L_0}\,. \tag{1.70}$$

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. (1.34). 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

$$\frac{L_s}{L_0} = \frac{\Delta E_1}{\Delta E_0}\,\frac{\Delta t_0}{\Delta t_1} = (1+z)^2\,. \tag{1.71}$$

From Eqs. (1.70) and (1.71) the luminosity distance reduces to

$$d_L = a_0 f_K(\chi)(1+z)\,. \tag{1.72}$$

Recall that the function $f_K(\chi)$ is given in Eq. (1.62) with the comoving distance

$$\chi = d_c = \frac{c}{a_0 H_0}\int_0^z \frac{\mathrm{d}\tilde{z}}{E(\tilde{z})}\,. \tag{1.73}$$

Then d_L can be expressed as

$$d_L = \frac{c(1+z)}{H_0\sqrt{\Omega_K^{(0)}}}\sinh\left(\sqrt{\Omega_K^{(0)}}\int_0^z \frac{\mathrm{d}\tilde{z}}{E(\tilde{z})}\right)\,, \tag{1.74}$$

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. (1.30) 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. (1.66), we find that d_L can be expanded around $z = 0$ as follows:

$$d_L = \frac{c}{H_0}\left[z + \left\{1 - \frac{E'(0)}{2}\right\} z^2 + \frac{1}{6}\left\{2E'(0)^2 - 3E'(0) - E''(0) + \Omega_K^{(0)}\right\} z^3 + \mathcal{O}(z^4)\right] . \tag{1.75}$$

In the small redshift region ($z \ll 1$) we have $d_L \simeq c\,z/H_0$. Using Eq. (1.35) we obtain

$$v \simeq H_0 d_L\,, \quad \text{for} \quad z \ll 1\,. \tag{1.76}$$

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

1.5.3 Angular diameter distance

The angular diameter distance d_A is defined by

$$d_A \equiv \frac{\Delta x}{\Delta\theta}\,, \tag{1.77}$$

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 (1.7) with (1.60) is given by

$$\Delta x = a(t_1) f_K(\chi)\Delta\theta\,. \tag{1.78}$$

Hence the diameter distance is

$$d_A = a(t_1) f_K(\chi) = \frac{a_0 f_K(\chi)}{1+z} = \frac{1}{1+z}\frac{c}{H_0\sqrt{\Omega_K^{(0)}}} \sinh\left(\sqrt{\Omega_K^{(0)}}\int_0^z \frac{\mathrm{d}\tilde{z}}{E(\tilde{z})}\right) , \tag{1.79}$$

where we have used $z = a_0/a(t_1) - 1$ and $f_K = c/(a_0H_0\sqrt{\Omega_K^{(0)}})\,\sinh(\sqrt{\Omega_K^{(0)}}\int_0^z \mathrm{d}\tilde{z}/E(\tilde{z}))$. Comparing Eq. (1.79) with Eq. (1.74), we notice the following relation

$$d_A = \frac{d_L}{(1+z)^2}\,. \tag{1.80}$$

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.

1.5.4 Degeneracy of the distance-redshift relation

All the distance definitions given above depend on the cosmological parameters through the integral $\chi = (c/(a_0H_0))\int_0^z \mathrm{d}\tilde{z}/E(\tilde{z})$ 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 ΛCDM 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. 1.1 we plot the lines of constant $f_K(\chi)$ in the plane $(\Omega_\Lambda^{(0)}, \Omega_m^{(0)})$ for redshifts that roughly corresponds 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. 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.

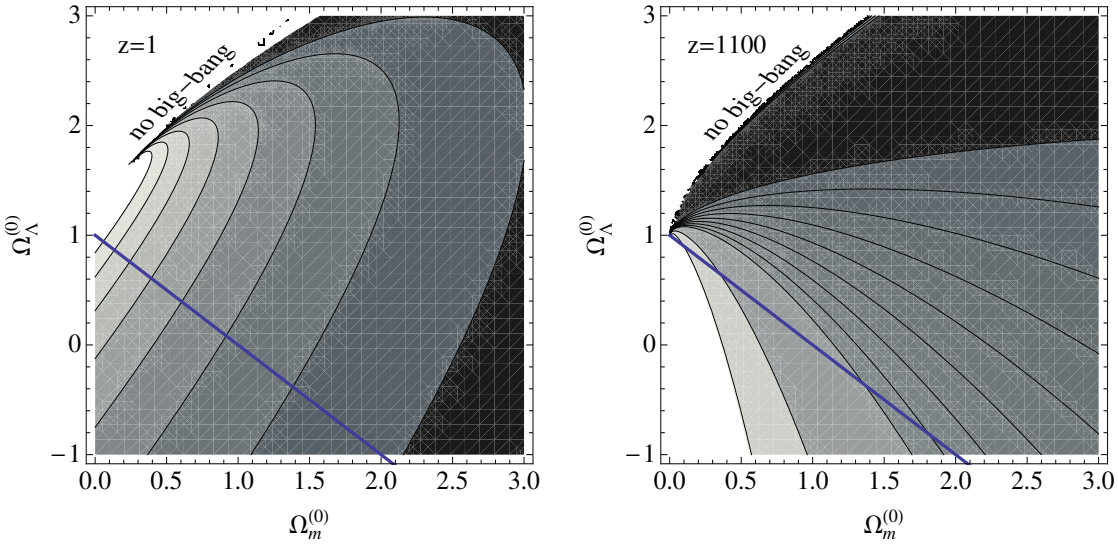

Figure 1.1: Contours of constant $f_K(\chi)$ for $z = 1$ and $z = 1100$. The straightl ine represents the flat model. In the left panel the contours range from 1400 to 2800 h^{-1} Mpc in steps of 200, from right to left. In the right panel they range from 300 to 5100 h^{-1} Mpc in steps of 600, from top to bottom (the last two contours are for 15000 and 30000 h^{-1} Mpc, respectively).

1.6 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 $\rho_{\rm DE}$ and pressure $P_{\rm DE}$). Since ρ_r and ρ_m evolves as $\rho_r \propto a^{-4}$ and $\rho_m \propto a^{-3}$, respectively, they can be expressed in the forms

$$\rho_r = \rho_r^{(0)}(a_0/a)^4 = \rho_r^{(0)}(1+z)^4\,, \tag{1.81}$$

$$\rho_m = \rho_m^{(0)}(a_0/a)^3 = \rho_m^{(0)}(1+z)^3\,. \tag{1.82}$$

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

$$1 + z_{\rm eq} = \frac{\rho_m^{(0)}}{\rho_r^{(0)}} = \frac{\Omega_m^{(0)}}{\Omega_r^{(0)}}\,, \tag{1.83}$$

where $\Omega_r^{(0)}$ is given by Eq. (1.53) with Eq. (1.51). 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. (1.57) and (1.58), respectively.

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

$$1 + z_{\rm eq} = 2.396 \times 10^4\, \Omega_m^{(0)} h^2\,. \tag{1.84}$$

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

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

$$\dot{\rho}_{\rm DE} + 3H(\rho_{\rm DE} + P_{\rm DE}) = 0\,. \tag{1.85}$$

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

$$\rho_{\rm DE} = \rho_{\rm DE}^{(0)} \exp\left[\int_0^z \frac{3(1+w_{\rm DE})}{1+\tilde{z}} d\tilde{z}\right]\,, \tag{1.86}$$

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

$$\rho_{\rm DE} = \rho_{\rm DE}^{(0)} (a/a_0)^{-3(1+\hat{w}_{\rm DE})}\,, \quad \hat{w}_{\rm DE}(z) = \frac{1}{\ln(1+z)} \int_0^z \frac{w_{\rm DE}(\tilde{z})}{1+\tilde{z}} {\rm d}\tilde{z}\,. \tag{1.87}$$

From the Friedmann equation (1.23) we have

$$H^2 = \frac{8\pi G}{3} (\rho_r + \rho_m + \rho_{\rm DE}) - \frac{K}{a^2}\,. \tag{1.88}$$

From Eq. (1.88) we see that the present density parameters defined in Eq. (1.30) obey the following relation

$$\Omega_r^{(0)} + \Omega_m^{(0)} + \Omega_{\rm DE}^{(0)} + \Omega_K^{(0)} = 1\,. \tag{1.89}$$

Then Eq. (1.88) can be written in the form

$$H^2(z) = H_0^2 \left[\Omega_r^{(0)} (1+z)^4 + \Omega_m^{(0)} (1+z)^3 + \Omega_{\rm DE}^{(0)} \exp\left\{ \int_0^z \frac{3(1+w_{\rm DE})}{1+\tilde{z}} {\rm d}\tilde{z} \right\} + \Omega_K^{(0)} (1+z)^2 \right]. \tag{1.90}$$

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

$$w_{\rm DE}(z) = \frac{(1+z)(E^2(z))' - 3E^2(z) - \Omega_r^{(0)}(1+z)^4 + \Omega_K^{(0)}(1+z)^2}{3\left[E^2(z) - \Omega_r^{(0)}(1+z)^4 - \Omega_m^{(0)}(1+z)^3 - \Omega_K^{(0)}(1+z)^2 \right]}\,, \tag{1.91}$$

where $E(z)$ is defined in Eq. (1.65) and a prime represents a derivative with respect to z. From Eq. (1.74) the quantity $E(z)$ can be written in terms of d_L:

$$E^2(z) = \frac{c^2(1+z)^2 \left[(1+z)^2 + \Omega_K^{(0)} H_0^2 d_L(z)^2 \right]}{[(1+z) H_0 d_L'(z) - H_0 d_L(z)]^2}\,. \tag{1.92}$$

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

$$E(z) = \frac{c}{H_0} \left[\frac{\rm d}{{\rm d}z} \left(\frac{d_L(z)}{1+z} \right) \right]^{-1}\,. \tag{1.93}$$

If the luminosity distance $d_L(z)$ is measured observationally, we can determine the evolution of $E(z)$ from Eq. (1.92) and hence $w_{\rm DE}(z)$ from Eqs. (1.91).

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. (1.91) reduces to

$$w_{\rm DE}(z) = \frac{(1+z)(E^2(z))' - 3E^2(z)}{3\left[E^2(z) - \Omega_m^{(0)}(1+z)^3 \right]}\,. \tag{1.94}$$

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 $w_{\rm DE}(z)$ can be negligible in the region $z \lesssim 1$, a small uncertainty on the curvature can produce a significant bias in $w_{\rm DE}$ in the high-redshift regime $z \gtrsim 1$ [81, 82]. In such a case we need to use the relation (1.91) rather than (1.94).

It is important to remark however that the basic observable quantity is $E(z)$, not $w_{\rm DE}(z)$. In fact, $w_{\rm DE}(z)$ cannot be determined entirely from $E(z)$, i.e. from measurements of the background expansion. From Eq. (1.94) 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. (1.94), 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

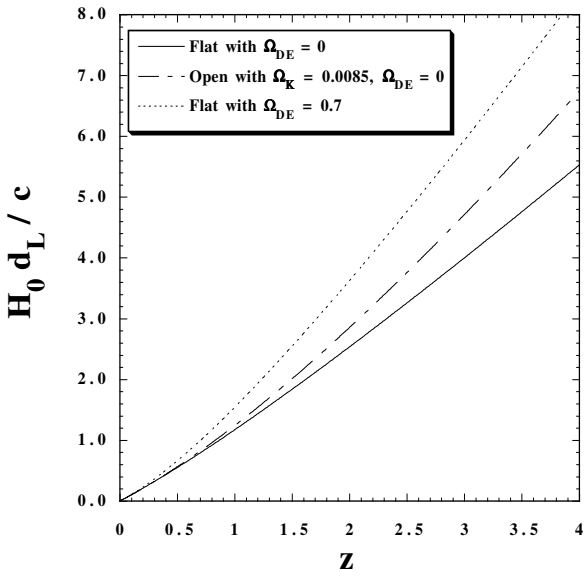

Figure 1.2: The luminosity distance d_L versus the redshift z for three cases: (a) a flat universe without dark energy, (b) an open universe ($\Omega_K^{(0)} = 0.0085$) without dark energy, and (c) a flat universe with the cosmological constant ($\Omega_{\rm DE}^{(0)} = 0.7$ and $w_{\rm DE} = -1$). The 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.

with dark energy. Of course if $w_{\rm DE}(z)$ 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 $-1.097 < w_{\rm DE} < -0.858$ at the 95 % confidence level [15]. Hence we cannot rule out the possibility that $w_{\rm DE}$ is smaller than -1. These cases are generally called "phantoms" or "ghosts" [83]. Since $P + \rho < 0$ in this case, Eq. (1.26) shows that ρ increases with time. From Eq. (1.23) the Hubble parameter H grows toward the future. When $w_{\rm DE}$ is constant and smaller than -1, the solution of the scale factor corresponding to the expanding universe is given by

$$a \propto (t_{\rm rip} - t)^{2/(3(1+w_{\rm DE}))}\,, \tag{1.95}$$

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

1.7 Accelerated expansion [b]

In 1998 Riess *et al.* and Perlmutter *et al.* 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 $d_L \simeq cz/H_0$, the absolute magnitude M is known. We take the value $h = 0.7$ for the Hubble constant. 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.

[b] Adapted from Amendola & Tsujikawa, *Dark Energy. Theory and Observations*, CUP 2010.

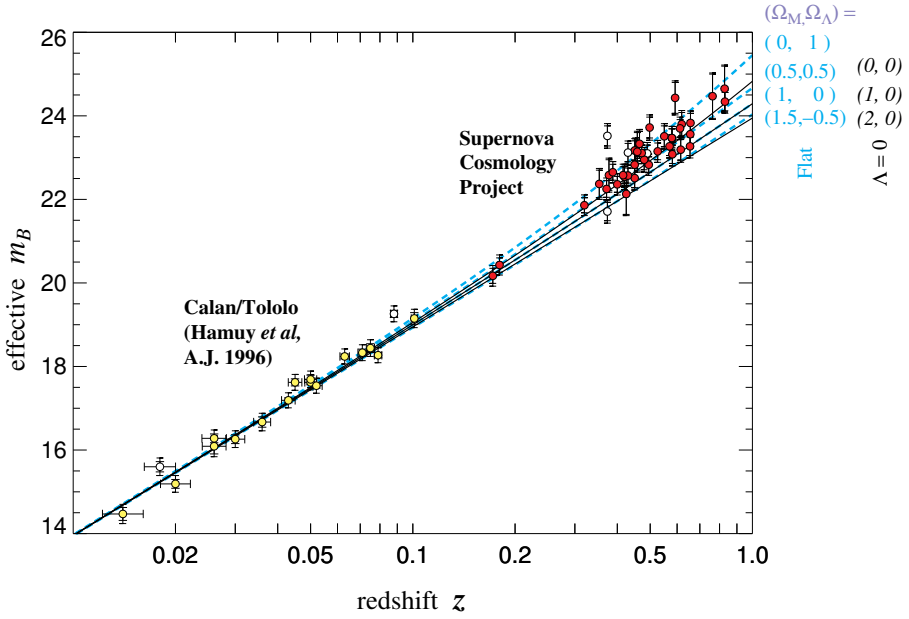

Figure 1.3: The effective apparent luminosity m_B versus the redshift z for 42 high-redshift SN Ia from the SCP and 18 low-redshift SN Ia from the Calan/Tololo Supernova Survey. The solid curves are the theoretical prediction for m_B for a number of cosmological models without the cosmological constant: $(\Omega_m^{(0)}, \Omega_\Lambda^{(0)}) = (0, 0)$ (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 Perlmutter et al. 1998.

Let us next use the high-redshift data reported by Perlmutter *et al.* . 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 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 $w_{\rm DE} = -1$ (i.e. the cosmological constant). Since $E(z) = [\Omega_m^{(0)}(1+z)^3 + \Omega_{\rm DE}^{(0)}]^{1/2}$ in this case, the luminosity distance reads

$$d_L(z) = \frac{c(1+z)}{H_0} \int_0^z \frac{{\rm d}\tilde{z}}{[(1-\Omega_{\rm DE}^{(0)})(1+\tilde{z})^3 + \Omega_{\rm DE}^{(0)}]^{1/2}}\,, \tag{1.96}$$

which can be evaluated numerically for given $\Omega_{\rm DE}^{(0)}$. In order to satisfy the observational data $(H_0/c)d_L(z = 0.657) = 0.920$ for 1997R, we require that $\Omega_{\rm DE}^{(0)} = 0.70$. Similarly we get $\Omega_{\rm DE}^{(0)} = 0.38$ 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 points 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, Perlmutter *et al.* (1998) 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. The apparent luminosity m gets larger for increasing luminosity distance d_L. Figure 1.3 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 $\Omega_\Lambda^{(0)} = 0$). This means that the SN are on average dimmer than expected in pure flat CDM. Dimmer sources imply, in turn, a larger luminosity distance and, therefore, a smaller value of H in the past: the conclusion is then that the Universe is accelerating.

From a full statistical 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 $\Omega_m^{(0)} = 0.28^{+0.09}_{-0.08}$ (1σ statistical) in the flat universe with the cosmological constant. After 1998 more SN Ia data have been collected by a number of high-redshift surveys–including SuperNova Legacy Survey (SNLS), Hubble Space Telescope (HST), and "Equation of State: SupErNovae trace Cosmic Expansion" (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. The ESSENCE project is a ground-based survey designed to detect about 200 SN Ia in the redshift range $z = 0.2$-0.8 to measure the equation of state of dark energy to better than 10 %. In Fig. 1.5 the observational contours on $(\Omega_m^{(0)}, w_{\rm DE})$ are plotted from the Union2 catalog of SN Ia by Amanullah *et al.* . 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 $w_{\rm DE}$, Fig. 1.5 clearly shows the presence of dark energy responsible for the late-time cosmic acceleration ($w_{\rm DE} < -1/3$). A modern publicly available catalog that includes most of the known supernovae is called JLA (Betoule et al. 2014, arXiv:1401.4064).

When combined with the results from CMB and from galaxy clustering on the plane Ω_Λ, Ω_m, the three probes intersect in a point $\Omega_m = 0.3, \Omega_\Lambda = 0.7$, which is denoted "concordance cosmology" (Fig. 1.6)

If the equation of state of dark energy varies in time, we need to parametrize $w_{\rm DE}$ as a function of the redshift z. A popular parametrization is the Chevalier-Polarski-Linder parametrization

$$w_{DE} = w_0 + w_a(1-a) = w_0 + w_a \frac{z}{1+z} \tag{1.97}$$

which can be thought as a Taylor expansion of an unknown $w_{DE}(a)$ around $a = 1$. The present value is then w_0 and the asymptotic future one is $w_0 + w_a$. Now from the conservation equation of ρ_{DE} we have

$$\dot{\rho}_{DE} + 3H\rho_{DE}(1 + w_{DE}(a)) = 0 \tag{1.98}$$

which can be integrated as

$$\int \frac{d\rho_{DE}}{\rho_{DE}} = -3\int \frac{da}{a}(1+w_{DE}) = -3\int \frac{da}{a}(1+w_0+w_a(1-a)) = -3(1+w_0+w_a)\log a + 3w_a(a-1) \tag{1.99}$$

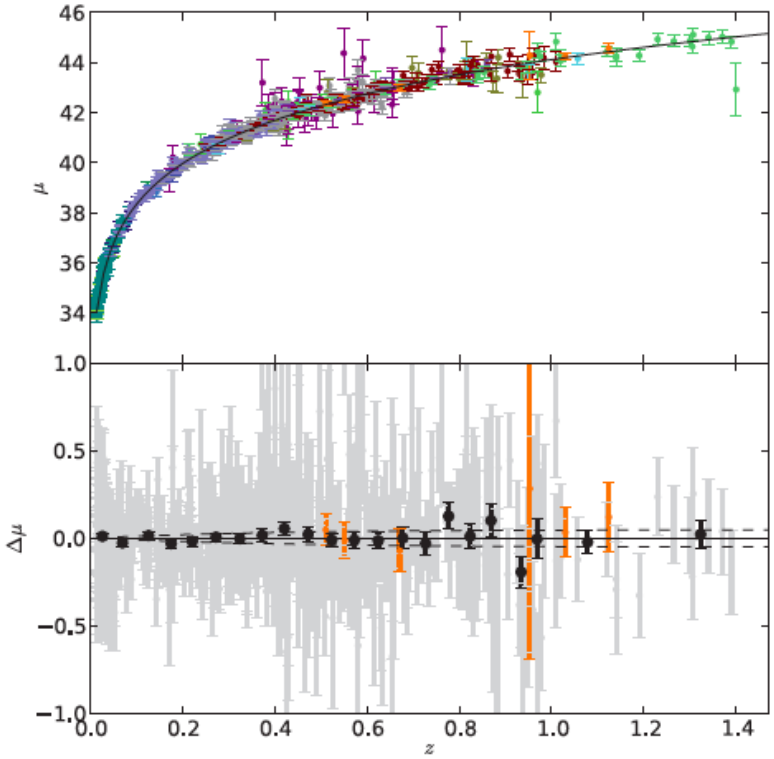

Figure 1.4: Union2 combined catalog of 557 SNIa. Here $\mu = 5 \log d_L + const$. From Amanullah et al. 2010 ApJ...716..712A *(Â© AAS. Reproduced with permission).* Bottom panel: residuals.

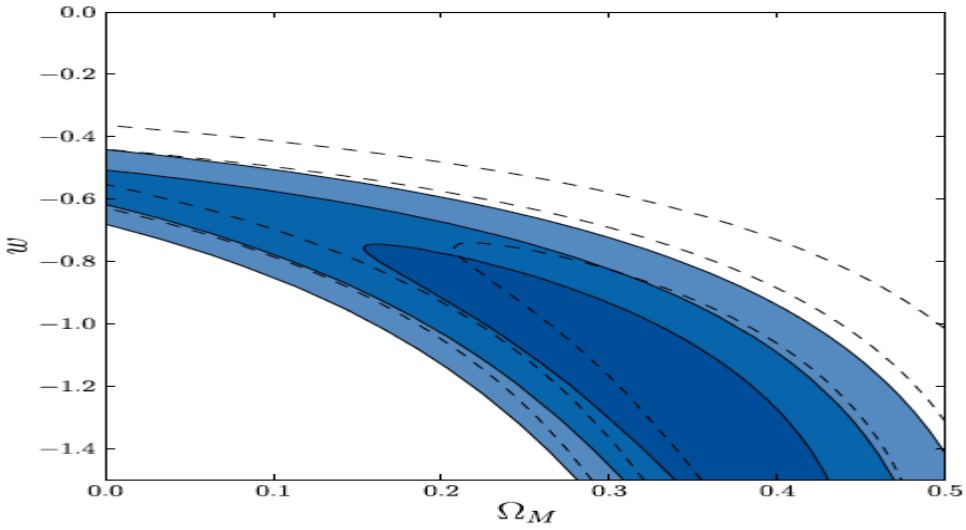

Figure 1.5: 68.3 %, 95.4 % and 99.7 % confidence level contours on $(\Omega_m^{(0)}, w_{\rm DE})$ from the SN Ia observations (denoted as Ω_M and w in the figure) compiled in Amanullah et al. 2010ApJ...716..712A *(Â© AAS. Reproduced with permission).*

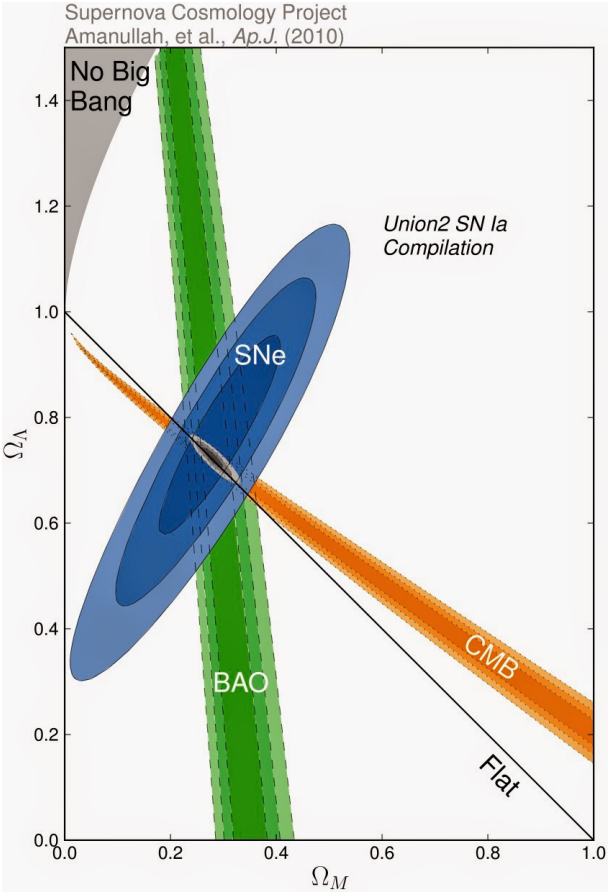

Figure 1.6: Combined results from supernovae, CMB and clustering of galaxies (BAO). The point where the three probes intersect is called "Concordance cosmology". (From Amanullah et al. 2010ApJ...716..712A, *Â© AAS. Reproduced with permission*).

so that

$$\rho_{DE} \sim a^{-3(1+w_0+w_a)} e^{3w_a(a-1)} \tag{1.100}$$

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 any function $w_{DE}(z)$ by using

$$H^2 = H_0^2(\Omega_{m0}(1+z)^3 + \Omega_{DE,0} f(z) + (1 - \Omega_{m0} - \Omega_{DE,0})(1+z)^2) \tag{1.101}$$

where, in general

$$f(z) = \exp \int_0^z \frac{3(1+w_{DE}(z))}{1+z} dz \tag{1.102}$$

However there is very little information on $w_{\rm DE}$ at $z > 0.5$ so $w_{DE}(z)$ cannot be well reconstructed beyond this. Any specific model of dark energy will in general provide a particular form of $w_{DE}(z)$ that will depend on the theoretical parameters.

Chapter 2

Cosmological constant[a]

The simplest candidate for dark energy is the cosmological constant Λ, which is called so because its energy density is constant in time and space. In fact the ΛCDM model has been systematically proved consistent with a large number of observations. The Lagrangian density for the ΛCDM model is simply given by the linear term in R plus Λ, see Eq. (2.2). 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.

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 ΛCDM 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 make it vanish or render 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 cosmological constant and dark energy are solved at the same time.

In this section we first present the action for the ΛCDM model. 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, observing vacuum energy in the laboratory, and the decoupling of Λ from gravity.

2.1 Einstein equations with the cosmological constant

The energy-momentum tensor $T_{\mu\nu}$ on the r.h.s. of the Einstein equations 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:

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}\,, \tag{2.1}$$

where Λ is the cosmological constant. It is interesting to note that these are the most general equations 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 (2.1) 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:

$$S = \frac{1}{16\pi G}\int \mathrm{d}^4x\sqrt{-g}\,(R - 2\Lambda) + S_m\,. \tag{2.2}$$

Repeating the steps we have already seen in the previous chapter, the variation of the action (2.2) with respect to $g^{\mu\nu}$ gives

$$\delta S = \frac{1}{16\pi G}\int \mathrm{d}^4x\left[\delta(\sqrt{-g})(g^{\mu\nu}R_{\mu\nu} - 2\Lambda) + \sqrt{-g}\,\delta g^{\mu\nu}R_{\mu\nu} + \sqrt{-g}\,g^{\mu\nu}\delta R_{\mu\nu}\right] + \delta S_m\,. \tag{2.3}$$

[a]Adapted from Amendola & Tsujikawa, *Dark Energy. Theory and Observations*, CUP 2010.

Since $\delta R_{\mu\nu} = (\delta\Gamma^\alpha_{\mu\nu})_{;\alpha} - (\delta\Gamma^\alpha_{\mu\alpha})_{;\nu}$ we have $g^{\mu\nu}\delta R_{\mu\nu} = (g^{\mu\nu}\delta\Gamma^\alpha_{\mu\nu} - g^{\mu\alpha}\delta\Gamma^\nu_{\mu\nu})_{;\alpha}$ and hence

$$\int \mathrm{d}^4x\sqrt{-g}\, g^{\mu\nu}\delta R_{\mu\nu} = \int \mathrm{d}^4x\sqrt{-g}(g^{\mu\nu}\delta\Gamma^\alpha_{\mu\nu} - g^{\mu\alpha}\delta\Gamma^\nu_{\mu\nu})_{;\alpha} = 0\,, \tag{2.4}$$

where we have employed Gauss's theorem in the last equality. This shows that the last term in the square bracket of Eq. (2.3) 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 first minor matrix, which does not depend on the element $g_{\mu\nu}$ itself, and then replacing $\mathcal{M}^{(\mu\nu)} = gg^{\mu\nu}$. Then Eq. (2.3) reads

$$\delta S = \frac{1}{16\pi G}\int \mathrm{d}^4x\sqrt{-g}\left(R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu}\right)\delta g^{\mu\nu} + \delta S_m\,. \tag{2.5}$$

The energy-momentum tensor $T_{\mu\nu}$ is defined from the variation of δS_m in terms of $g^{\mu\nu}$:

$$\delta S_m = -\frac{1}{2}\int \mathrm{d}^4x\sqrt{-g}\, T_{\mu\nu}\delta g^{\mu\nu}\,. \tag{2.6}$$

Then Eq. (2.5) reduces to

$$\delta S = \frac{1}{16\pi G}\int \mathrm{d}^4x\sqrt{-g}\left(R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} - 8\pi G T_{\mu\nu}\right)\delta g^{\mu\nu}\,. \tag{2.7}$$

The Einstein equations (2.1) follow from the action principle, $\delta S = 0$.

2.2 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 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 (1.7) the Einstein equations (2.1) read

$$H^2 = \frac{8\pi G}{3}\rho - \frac{K}{a^2} + \frac{\Lambda}{3}\,, \tag{2.8}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3}\,. \tag{2.9}$$

From Eq. (2.9) 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

$$\rho = \frac{\Lambda}{4\pi G}\,, \qquad \frac{K}{a^2} = \Lambda\,. \tag{2.10}$$

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 LemaÃ®tre. In fact, if $\Lambda/3 > (4\pi G\rho)/3$, Eq. (2.9) shows that the universe departs from the static point given in Eq. (2.10) 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 LemaÃ®tre [131] studied the relation between the observed results of the redshift and the homogeneous universe dominated by a pressureless dust. In LemaÃ®tre'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 ($a \propto$ 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.

LemaÃ®tre's model is the first "hot Big Bang" model, in which the matter density ρ goes to infinity as $a \to 0$. Apart from the existence of the period (ii) LemaÃ®tre'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 LemaÃ®tre, influenced by de Sitter's accelerated solution, produced the first consistent dark energy model.

In 1929 Hubble formulated Hubble's law, today called Hubble-LemaÃ®tre's law, (1.38) 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.

2.3 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. (2.8)]:

$$\Lambda \approx H_0^2 = (2.1332h \times 10^{-42}\ \mathrm{GeV})^2\,. \tag{2.11}$$

If we interpret this as an energy density, it is equivalent to

$$\rho_\Lambda \approx \frac{\Lambda m_{\rm pl}^2}{8\pi} \approx 10^{-47}\ \mathrm{GeV}^4 \approx 10^{-123} m_{\rm pl}^4\,, \tag{2.12}$$

where we have used $h \approx 0.7$ and $m_{\rm pl} \approx 10^{19}$ GeV.

Suppose that the energy density (2.12) 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 $k_{\rm max}$ ($\gg m$), we obtain the vacuum energy density

$$\rho_{\rm vac} = \int_0^{k_{\rm max}} \frac{{\rm d}^3k}{(2\pi)^3}\frac{1}{2}\sqrt{k^2+m^2}\,. \tag{2.13}$$

Since the integral is dominated by the mode with large k with $k \gg m$, we find that

$$\rho_{\rm vac} = \int_0^{k_{\rm max}} \frac{4\pi k^2 {\rm d}k}{(2\pi)^3}\frac{1}{2}\sqrt{k^2+m^2} \approx \frac{k_{\rm max}^4}{16\pi^2}\,. \tag{2.14}$$

General Relativity is believed to be valid up to the Planck scale $m_{\rm pl}$. Taking the cut-off scale $k_{\rm max}$ to be $m_{\rm pl}$, the vacuum energy density can be estimated as

$$\rho_{\rm vac} \simeq 10^{74}\ \mathrm{GeV}^4\,. \tag{2.15}$$

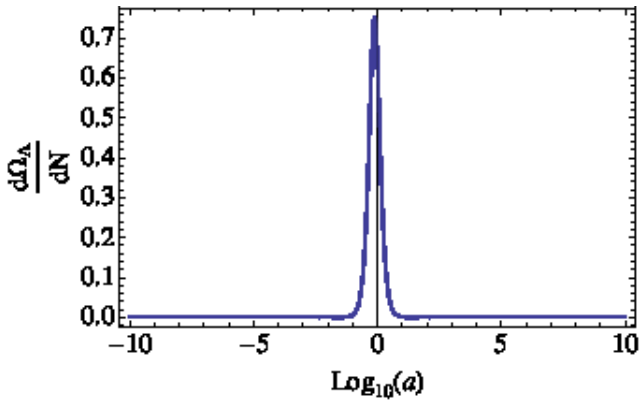

Figure 2.1: Plot of $\mathrm{d}\Omega_\Lambda/\mathrm{d}N$ versus $\log_{10} a$, assuming flat space with $\Omega_\Lambda^{(0)} = 0.7$. The spike is very close to the present epoch: this is the coincidence problem.

This is about 10^{121} times larger than the observed value (2.12). Note that this situation is not improved much by taking other energy scales appearing in particle physics. For the QCD scale $k_{\max} \approx 0.1$ GeV we have $\rho_{\rm vac} \approx 10^{-3}$ 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 counter part 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]. However it is known that supersymmetry is broken at sufficient high energies (around the scale $M_{\rm SUSY} = 10^3$ GeV if it is relevant to the hierarchy problem of gravitational 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.

2.4 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

$$z_{\rm coinc} = \left(\frac{\Omega_\Lambda^{(0)}}{1-\Omega_\Lambda^{(0)}}\right)^{1/3} - 1\,, \tag{2.16}$$

which, for $\Omega_\Lambda^{(0)} = 0.7$, amounts to $z_{\rm coinc} \approx 0.3$. This problem is called the *coincidence problem*.

To illustrate the issue, we plot the evolution of the derivative $\mathrm{d}\Omega_\Lambda/\mathrm{d}N$ (where $N = \ln a$) in Fig. 2.1. We find that the only epoch in which this quantity is not close to zero is the present. 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 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 $z_{\rm coinc}$ also turn out to be very close to zero. Therefore we discuss this problem in terms of a general dark energy density $\rho_{\rm DE}$.

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 $\rho_{\rm DE}$ responds to the trend of ρ_m and catches up with it irrespective of the initial conditions of $\rho_{\rm DE}$. In this case $\Omega_{\rm DE}$ 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 Sec. 3.2.3.

The second class of explanation argues that there is no coincidence and in fact Ω_m and $\Omega_{\rm DE}$ 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. 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 $\rho_{\rm DE} \approx \rho_m$ 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.

The fourth class is the "backreaction" argument. The coincidence between $\rho_{\rm DE}$ 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. 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 LemaÃ®tre-Tolman-Bondi void, the acceleration of the recession rate between nearby and distant sources becomes a distance-dependent, but practically always decelerated, Hubble rate.

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.

Chapter 3

Dark energy models[a]

If the cosmological constant problem is solved in a way that Λ completely vanishes, we need to find out alternative models of dark energy. As we already mentioned in 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 put 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 [180]. 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.

Caldwell *et al.* [35] named quintessence a canonical scalar field ϕ with a potential $V(\phi)$ responsible for the late-time cosmic acceleration[b]. Unlike the cosmological 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 have 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 $\rho_{\rm DE} \simeq 10^{-123} m_{\rm pl}^4$ and a mass scale $m_\phi \lesssim 10^{-33}\,{\rm eV}$. 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.

[a] Adapted from Amendola & Tsujikawa, *Dark Energy. Theory and Observations*, CUP 2010.

[b] 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.

3.1 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

$$S = \int \mathrm{d}^4x\sqrt{-g}\left[\frac{1}{2\kappa^2}R + \mathcal{L}_\phi\right] + S_M\,, \qquad \mathcal{L}_\phi = -\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi)\,, \tag{3.1}$$

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 (1.26), i.e.

$$\dot{\rho}_M + 3H(\rho_M + P_M) = 0\,. \tag{3.2}$$

The energy-momentum tensor of quintessence is [73]

$$T^{(\phi)}_{\mu\nu} = -\frac{2}{\sqrt{-g}}\frac{\delta(\sqrt{-g}\mathcal{L}_\phi)}{\delta g^{\mu\nu}} \tag{3.3}$$

$$= \partial_\mu\phi\partial_\nu\phi - g_{\mu\nu}\left[\frac{1}{2}g^{\alpha\beta}\partial_\alpha\phi\partial_\beta\phi + V(\phi)\right]. \tag{3.4}$$

In a FLRW background, the energy density ρ_ϕ and the pressure P_ϕ of the field are

$$\rho_\phi = -T^{0\,(\phi)}_0 = \frac{1}{2}\dot{\phi}^2 + V(\phi)\,, \quad P_\phi = \frac{1}{3}T^{i\,(\phi)}_i = \frac{1}{2}\dot{\phi}^2 - V(\phi)\,, \tag{3.5}$$

which gives the equation of state

$$w_\phi \equiv \frac{P_\phi}{\rho_\phi} = \frac{\dot{\phi}^2 - 2V(\phi)}{\dot{\phi}^2 + 2V(\phi)}\,. \tag{3.6}$$

In the flat universe ($K = 0$) the following equations of motion follow from Eqs. (1.23) and (1.24):

$$H^2 = \frac{\kappa^2}{3}\left[\frac{1}{2}\dot{\phi}^2 + V(\phi) + \rho_M\right], \tag{3.7}$$

$$\dot{H} = -\frac{\kappa^2}{2}\left(\dot{\phi}^2 + \rho_M + P_M\right), \tag{3.8}$$

where $\kappa^2 = 8\pi G$. The variation of the action (3.1) with respect to ϕ gives

$$\ddot{\phi} + 3H\dot{\phi} + V_{,\phi} = 0\,, \tag{3.9}$$

where $V_{,\phi} \equiv \mathrm{d}V/\mathrm{d}\phi$. The Klein-Gordon equation (3.9) can be also derived by using the continuity equation $\dot{\rho}_\phi + 3H(\rho_\phi + P_\phi) = 0$ or by combining Eqs. (3.2), (3.7) and (3.8).

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. (3.6). 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]

$$\epsilon_s \equiv \frac{1}{2\kappa^2}\left(\frac{V_{,\phi}}{V}\right)^2, \qquad \eta_s \equiv \frac{V_{,\phi\phi}}{\kappa^2 V}\,. \tag{3.10}$$

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. (3.7) and (3.9).

From Eq. (3.9) the deviation of w_ϕ from -1 is given by

$$1 + w_\phi = \frac{V_{,\phi}^2}{9H^2(\xi_s + 1)^2 \rho_\phi}\,, \tag{3.11}$$

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. (3.7), 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 [181]. In the 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

- $V(\phi) = M^{4+n}\phi^{-n} \quad (n > 0)\,,$
- $V(\phi) = M^{4+n}\phi^{-n} \exp(\alpha\phi^2/m_{\rm pl}^2)\,.$

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 the 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 $V_0 = 0$) [182], while the mass scale M is very different. Note that the model with $n = 1$ was originally proposed in Ref. [183] to replace the cosmological constant by a slowly varying field and was revised in Ref. [184] 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, i.e. when

$$m_\phi \equiv \left(\frac{\partial^2 V}{\partial \phi^2}\right)^{1/2} \simeq H_0 \tag{3.12}$$

3.2 Dynamical system approach

In order to study cosmological dynamics in the presence of a scalar field and a background fluid with EoS w_M (which could be 0 for matter or 1/3 for radiation), it is convenient to introduce the following dimensionless variables:

$$x_1 \equiv \frac{\kappa\dot{\phi}}{\sqrt{6}H}\,, \qquad x_2 \equiv \frac{\kappa\sqrt{V}}{\sqrt{3}H}\,. \tag{3.13}$$

Then Eq. (3.7) can be written as

$$\Omega_M \equiv \frac{\kappa^2 \rho_M}{3H^2} = 1 - x_1^2 - x_2^2 \,. \tag{3.14}$$

We also define the energy fraction of dark energy

$$\Omega_\phi \equiv \frac{\kappa^2 \rho_\phi}{3H^2} = x_1^2 + x_2^2 \,, \tag{3.15}$$

which satisfies the relation $\Omega_M + \Omega_\phi = 1$. From Eq. (3.8) we obtain

$$\frac{\dot{H}}{H^2} = -3x_1^2 - \frac{3}{2}(1 + w_M)(1 - x_1^2 - x_2^2) = -\frac{3}{2}(1 + w_{\rm eff}), \tag{3.16}$$

where we define an effective equation of state $w_{\rm eff}$ (not necessarily a constant) for a model with several components as

$$w_{\rm eff} = \frac{p_{\rm tot}}{\rho_{\rm tot}} = \sum_i w_i \Omega_i \tag{3.17}$$

In the present case, the effective equation of state is given by

$$w_{\rm eff} = w_M + (1 - w_M)x_1^2 - (1 + w_M)x_2^2 \,. \tag{3.18}$$

Clearly, since the total energy density is conserved and therefore obeys the equation $\ddot{a}/a = -(4\pi/3)\rho_{tot}(1+3w_{\rm eff})$, one has that if $w_{\rm eff} < -1/3$ the expansion is accelerated. The equation of state of dark energy (3.6) can be expressed as

$$w_\phi = \frac{x_1^2 - x_2^2}{x_1^2 + x_2^2} \,. \tag{3.19}$$

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. (3.9) and (3.16), we obtain the following equations

$$\frac{{\rm d}x_1}{{\rm d}N} = -3x_1 + \frac{\sqrt{6}}{2}\lambda x_2^2 + \frac{3}{2}x_1 \left[(1 - w_M)x_1^2 + (1 + w_M)(1 - x_2^2)\right] \,, \tag{3.20}$$

$$\frac{{\rm d}x_2}{{\rm d}N} = -\frac{\sqrt{6}}{2}\lambda x_1 x_2 + \frac{3}{2}x_2 \left[(1 - w_M)x_1^2 + (1 + w_M)(1 - x_2^2)\right] \,, \tag{3.21}$$

where

$$\lambda \equiv -\frac{V_{,\phi}}{\kappa V} \,. \tag{3.22}$$

The quantity λ characterizes the slope of the field potential, which obeys the following equation

$$\frac{{\rm d}\lambda}{{\rm d}N} = -\sqrt{6}\lambda^2(\Gamma - 1)x_1 \,, \tag{3.23}$$

where

$$\Gamma \equiv \frac{V V_{,\phi\phi}}{V_{,\phi}^2} \,. \tag{3.24}$$

If λ is constant, the integration of Eq. (3.22) yields an exponential potential

$$V(\phi) = V_0 e^{-\kappa\lambda\phi} \,. \tag{3.25}$$

From Eq. (3.24) this potential corresponds to $\Gamma = 1$. In this case the autonomous equations (3.20) and (3.21) 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. (3.23) as well as Eqs. (3.20) and (3.21). 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. (3.23). 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. (3.16) and (3.21) 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. (3.25) and then proceed to the case of non-constant λ.

3.2.1 Exponential potential

We can derive fixed points of the system by setting $\mathrm{d}x_1/\mathrm{d}N = \mathrm{d}x_2/\mathrm{d}N = 0$ in Eqs. (3.20) and (3.21). 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) $(x_1, x_2) = (0, 0), \quad \Omega_\phi = 0, \quad w_{\rm eff} = w_M, \quad w_\phi$: undefined.
- (b1) $(x_1, x_2) = (+1, 0), \quad \Omega_\phi = 1, \quad w_{\rm eff} = 1, \quad w_\phi = 1.$
- (b2) $(x_1, x_2) = (-1, 0), \quad \Omega_\phi = 1, \quad w_{\rm eff} = 1, \quad w_\phi = 1.$
- (c) $(x_1, x_2) = (\lambda/\sqrt{6}, [1 - \lambda^2/6]^{1/2}), \quad \Omega_\phi = 1, \quad w_{\rm eff} = -1 + \lambda^2/3, \quad w_\phi = -1 + \lambda^2/3.$
- (d) $(x_1, x_2) = (\sqrt{3/2}(1+w_M)/\lambda, [3(1-w_M^2)/2\lambda^2]^{1/2}), \quad \Omega_\phi = 3(1+w_M)/\lambda^2, \quad w_{\rm eff} = w_M, \qquad w_\phi = w_M.$

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 $w_{\rm eff} < -1/3$, i.e. $\lambda^2 < 2$. In the limit that $\lambda \to 0$ (i.e. $V(\phi) \to V_0$) we recover the equation of state of cosmological constant ($w_{\rm eff} = w_\phi = -1$). 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:

$$x_1 = x_1^{(c)} + \delta x_1\,, \quad x_2 = x_2^{(c)} + \delta x_2\,. \tag{3.26}$$

Linearizing Eqs. (3.20) and (3.21) leads to the first-order differential equations

$$\frac{\mathrm{d}}{\mathrm{d}N}\begin{pmatrix} \delta x_1 \\ \delta x_2 \end{pmatrix} = \mathcal{M}\begin{pmatrix} \delta x_1 \\ \delta x_2 \end{pmatrix}\,, \tag{3.27}$$

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

$$\mu_{1,2} = \frac{1}{2}\left[a_{11} + a_{22} \pm \sqrt{\mathcal{D}}\right]\,, \tag{3.28}$$

where a_{ij} are the components of the matrix and

$$\mathcal{D} \equiv (a_{11} + a_{22})^2 - 4(a_{11}a_{22} - a_{12}a_{21}) \tag{3.29}$$

. The general linearized solution around each fixed point can be written then as

$$x_i = x_i^{(c)} + \alpha_{i\,1}e^{\mu_1 N} + \alpha_{i\,2}e^{\mu_2 N}\,, \tag{3.30}$$

where $\alpha_{i\,1}$ and $\alpha_{i\,2}$ are coefficients obtained from the eigenvectors. The eigenvalues determine therefore the behavior of solutions near the fixed points.

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 matrix $\mathcal{M}$ is singular, 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.

We now restrict our attention to $w_M = 0$. The eigenvalues of the above fixed points are given by (see the problem [7.1])

- Point (a): $\mu_1 = -\frac{3}{2}, \quad \mu_2 = \frac{3}{2}$.
- Point (b1): $\mu_1 = 3 - \frac{\sqrt{6}}{2}\lambda, \quad \mu_2 = 3$.
- Point (b2): $\mu_1 = 3 + \frac{\sqrt{6}}{2}\lambda, \quad \mu_2 = 3$.
- Point (c): $\mu_1 = \frac{1}{2}(\lambda^2 - 6), \quad \mu_2 = \lambda^2 - 3$.
- Point (d): $\mu_{1,2} = -\frac{3}{4}\left[1 \pm \sqrt{1 - \frac{8(\lambda^2-3)}{\lambda^2}}\right]$.

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 $\lambda^2 < 3$ and saddle point for $3 < \lambda^2 < 6$.
- Point (d): Saddle for $\lambda^2 < 3$, stable node for $3 < \lambda^2 < \frac{24}{7}$ and stable spiral for $\lambda^2 > \frac{24}{7}$.

A matter dominated epoch can be realized either by the point (a) or (d). For any λ, there is only one final stable point, either (c) or (d). When $\lambda^2 > 3$ 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 (Ω_ϕ =constant) to connect to the accelerated epoch. In order to give rise to tracking behavior in which Ω_ϕ evolves to catch 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 Sec. 3.2.2. 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. 3.1 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. (3.14), 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 $\lambda^2 < 3$.

3.2.2 Other potentials

If λ is not constant, we need to solve Eq. (3.23) 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 [185, 186], provided that the time scale for the variation of λ is much less than H^{-1}.

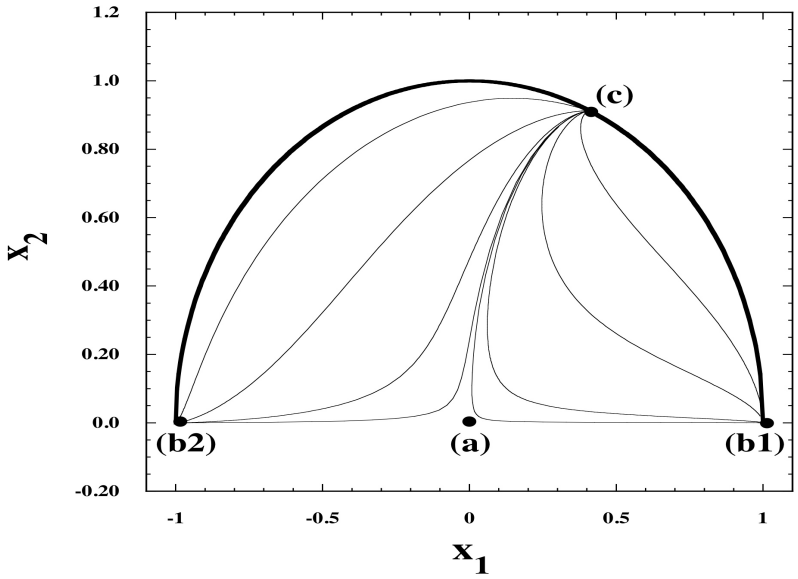

Figure 3.1: The trajectories of solutions for the exponential potential (3.25) with model parameters $\lambda = 1$ and $w_M = 0$. In this case the attractor is the accelerated point (c) $(x_1, x_2) = (0.4082, 0.9129)$. The matter point (a) is a saddle, whereas the points (b1) and (b2) are unstable nodes. The thick curve is the border of the allowed region characterized by $x_2 = \sqrt{1 - x_1^2}$.

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

$$\Gamma = \frac{V V_{,\phi\phi}}{V_{,\phi}^2} > 1\,, \tag{3.31}$$

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 (3.31) is the so-called tracking condition under which the field density eventually catches up that of the background fluid.

The condition (3.31) can be also derived in the following way [39]. We first define the following quantity

$$x \equiv \frac{1 + w_\phi}{1 - w_\phi} = \frac{\dot{\phi}^2}{2V}\,. \tag{3.32}$$

Taking the derivative of x in terms of N and using the definition (3.15), we find

$$\frac{V_{,\phi}}{\kappa V} = \pm\sqrt{\frac{3(1 + w_\phi)}{\Omega_\phi}}\left(1 + \frac{1}{6}\frac{\mathrm{d}\,\ln x}{\mathrm{d}N}\right)\,, \tag{3.33}$$

where the plus and minus signs correspond to the cases $\dot{\phi} < 0$ and $\dot{\phi} > 0$, respectively. Differentiating Eq. (3.33) with respect to ϕ, we get the following relation:

$$\Gamma = 1 + \frac{3(1 - \Omega_\phi)(w_M - w_\phi)}{(1 + w_\phi)(6 + y')} - \frac{y'}{(1 + w_\phi)(6 + y')(1 + x)} - \frac{2y''}{(1 + w_\phi)(6 + y')^2}\,, \tag{3.34}$$

where $y' \equiv \mathrm{d}\,\ln x/\mathrm{d}N = 2(dw_\phi/dN)/(w^2 - 1)$. 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. (3.34) 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:

$$w_\phi \simeq \frac{w_M - 2(\Gamma - 1)}{1 + 2(\Gamma - 1)}\,. \tag{3.35}$$

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 (3.31) for Γ nearly constant ($|\mathrm{d}(\Gamma-1)/\mathrm{d}N| \ll |\Gamma-1|$).

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 (3.31) 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.

$$\phi > \frac{n}{4\sqrt{\pi}} m_{\rm pl}\,, \tag{3.36}$$

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 (3.7) one can estimate the present potential energy of quintessence to be $V(\phi_0) \approx H_0^2 m_{\rm pl}^2$, where $\phi_0 \approx m_{\rm pl}$. Then the mass M is constrained to be

$$M \approx \left(\frac{H_0}{m_{\rm pl}}\right)^{\frac{2}{4+n}} m_{\rm pl} \approx 10^{-\frac{46-19n}{4+n}}\ {\rm GeV}\,, \tag{3.37}$$

where we have used $H_0 \approx 10^{-42}\,{\rm GeV}$. For $n = 2$ and $n = 4$ we have that $M \approx 10^{-1}$ GeV and $M \approx 10^4$ 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 (3.31). Since $|\lambda| = (n/\sqrt{8\pi})(m_{\rm pl}/|\phi|)$, the late-time acceleration occurs only in the region $|\phi| > n m_{\rm pl}/4\sqrt{\pi}$. 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 $|\phi_i| \gtrsim m_{\rm pl}$ 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 $m_{\rm pl}$. 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.

3.2.3 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 details. To be concrete we consider the inverse power-law potential

$$V(\phi) = M^{4+n}\phi^{-n}\,, \quad (n > 0)\,. \tag{3.38}$$

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. (3.7) and (3.8) are given by $\rho_M = \rho_r + \rho_m$ and $P_M = \rho_r/3$, respectively. In addition to the variable x_1 and x_2 defined in Eq. (3.13) 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

$$\Omega_\phi = x_1^2 + x_2^2\,, \quad \Omega_r = x_3^2\,, \quad \Omega_m = 1 - x_1^2 - x_2^2 - x_3^2\,. \tag{3.39}$$

The effective equation of state reads

$$w_{\rm eff} = x_1^2 - x_2^2 + x_3^2/3\,. \tag{3.40}$$

The equation of state of quintessence is the same as Eq. (3.6).

The equations for x_1, x_2, and x_3 are

$$\frac{\mathrm{d}x_1}{\mathrm{d}N} = -3x_1 + \frac{\sqrt{6}}{2}\lambda x_2^2 + \frac{1}{2}x_1(3 + 3x_1^2 - 3x_2^2 + x_3^2)\,, \tag{3.41}$$

$$\frac{\mathrm{d}x_2}{\mathrm{d}N} = -\frac{\sqrt{6}}{2}\lambda x_1 x_2 + \frac{1}{2}x_2(3 + 3x_1^2 - 3x_2^2 + x_3^2)\,, \tag{3.42}$$

$$\frac{\mathrm{d}x_3}{\mathrm{d}N} = -2x_3 + \frac{1}{2}x_3(3 + 3x_1^2 - 3x_2^2 + x_3^2)\,. \tag{3.43}$$

Using the fact that $\Gamma = (n+1)/n$ for the potential (3.38), the equation for λ reads

$$\frac{\mathrm{d}\lambda}{\mathrm{d}N} = -\sqrt{6}\frac{\lambda^2}{n}x_1\,. \tag{3.44}$$

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. (3.44) is negative so that λ decreases with time. The equations (3.41)-(3.44) are the autonomous equations to be solved numerically.

From Eq. (3.35) the equation of state of quintessence in the tracking regime is given by

$$w_\phi \simeq \frac{nw_M - 2}{n+2}\,. \tag{3.45}$$

If the tracking occurs during the matter-dominated epoch ($w_M = 0$), then $w_\phi \simeq -2/(n+2)$. From Eq. (3.33) the following relation holds

$$\frac{1}{6}\frac{\mathrm{d}\,\ln x}{\mathrm{d}N} = \Delta(t) - 1\,, \quad \text{where} \quad \Delta(t) \equiv \lambda\sqrt{\frac{\Omega_\phi}{3(1+w_\phi)}}\,. \tag{3.46}$$

From the definition of x in Eq. (3.32) we also obtain

$$\frac{1}{6}\frac{\mathrm{d}\,\ln x}{\mathrm{d}N} = \frac{1}{3(1-w_\phi^2)}\frac{\mathrm{d}w_\phi}{\mathrm{d}N}\,. \tag{3.47}$$

Since w_ϕ is nearly constant for tracker solutions, it follows from Eqs. (3.46) and (3.47) that $\Delta \simeq 1$. Hence the tracker solution is characterized by

$$\Omega_\phi \simeq \frac{3(1+w_\phi)}{\lambda^2}\,, \tag{3.48}$$

where w_ϕ is given in Eq. (3.45). 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 (3.48) 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 (3.48) 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.

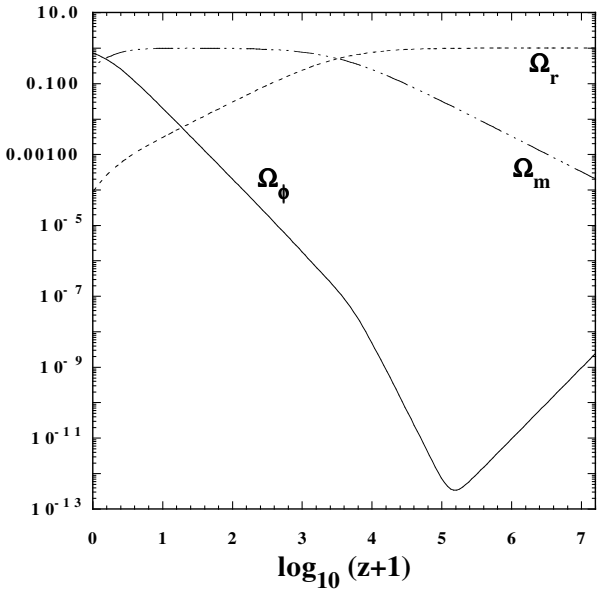

Figure 3.2: Evolution of Ω_ϕ, Ω_m, and Ω_r for the model $V(\phi) = M^5\phi^{-1}$ versus the redshift z. Initially Ω_ϕ rapidly decreases because the field equation of state is given by $w_\phi \simeq 1$. This is followed by the phase with a nearly frozen scalar field so that Ω_ϕ begins to grow. Finally 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 $\log_{10}(z+1) = 7.21$.

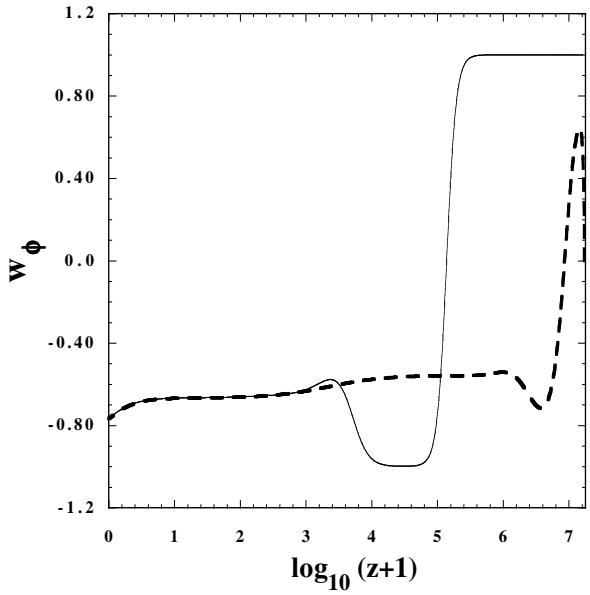

Figure 3.3: Evolution of w_ϕ for the model $V(\phi) = M^5\phi^{-1}$ with two different initial conditions. The solid curve corresponds to the case shown in Fig. 3.2. The dashed curve corresponds to the case with initial conditions $x_1 = 1.0 \times 10^{-8}$, $x_2 = 1.0 \times 10^{-8}$, $x_3 = 0.9999$, 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 ($10 \lesssim z \lesssim 10^3$).

Chapter 4

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, 270, 40, 271], $f(R)$ gravity [272], and dilaton gravity [210] 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 $H\delta\rho_m$.

4.1 Coupled quintessence with an exponential potential

Let us consider an interaction between a scalar field ϕ and a non-relativistic matter in the form

$$\nabla_\mu T^\mu_{\nu(\phi)} = -Q\, T_M \nabla_\nu \phi\,, \qquad \nabla_\mu T^\mu_{\nu(M)} = +Q\, T_M \nabla_\nu \phi\,, \tag{4.1}$$

where $T^\mu_{\nu(\phi)}$ and $T^\mu_{\nu(M)}$ are the energy momentum tensors of ϕ and non-relativistic matter, respectively, with a trace $T_M = -\rho_M + 3P_M$ of the matter. Since the radiation is traceless ($\rho_M = 3P_M$), the coupling-dependent terms vanish in Eq. (4.1). 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.

Although we assume the baryons to be uncoupled, for as concerns cosmology 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 later on. In this section we shall use the unit $\kappa^2 = 1$ unless otherwise stated.

The field Lagrangian density of the coupled quintessence is $\mathcal{L}_\phi = -(1/2)g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi) + \mathcal{L}_{\rm int}$, where the part $\mathcal{L}_{\rm int}$ gives rise to the interacting energy-momentum tensor given in Eq. (4.1) [a]. For the field potential $V(\phi)$ we can take the exponential type

$$V(\phi) = V_0 e^{-\lambda\phi}\,, \tag{4.2}$$

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. (4.1), the field ϕ, non-relativistic matter, and radiation obey the following equations of motion, respectively, in the flat FLRW background:

$$\dot{\rho}_\phi + 3H(\rho_\phi + P_\phi) = -Q\rho_m\dot{\phi}\,, \tag{4.3}$$
$$\dot{\rho}_m + 3H\rho_m = Q\rho_m\dot{\phi}\,, \tag{4.4}$$
$$\dot{\rho}_r + 4H\rho_r = 0\,, \tag{4.5}$$

together with the usual Friedmann equation

$$3H^2 = \rho_\phi + \rho_m + \rho_r\,. \tag{4.6}$$

Since $\rho_\phi = (1/2)\dot{\phi}^2 + V(\phi)$ and $P_\phi = (1/2)\dot{\phi}^2 - V(\phi)$, Eq. (4.3) can be written as

$$\ddot{\phi} + 3H\dot{\phi} + V_{,\phi} = -Q\rho_m\,. \tag{4.7}$$

In order to study the dynamics of the system we introduce the following variables

$$x_1 \equiv \frac{\dot{\phi}}{\sqrt{6}H}\,, \quad x_2 \equiv \frac{\sqrt{V}}{\sqrt{3}H}\,, \quad x_3 \equiv \frac{\sqrt{\rho_r}}{\sqrt{3}H}\,. \tag{4.8}$$

Taking the derivative of Eq. (4.6) in terms of the number of e-foldings N together with the use of Eqs. (4.3)-(4.5), we obtain

$$\frac{1}{H}\frac{{\rm d}H}{{\rm d}N} = -\frac{1}{2}\left(3 + 3x_1^2 - 3x_2^2 + x_3^2\right)\,. \tag{4.9}$$

The effective equation of state is

$$w_{\rm eff} = x_1^2 - x_2^2 + x_3^2/3\,. \tag{4.10}$$

The equation of state of the scalar field w_ϕ and the density parameter Ω_ϕ are

$$w_\phi = \frac{x_1^2 - x_2^2}{x_1^2 + x_2^2}\,, \qquad \Omega_\phi = x_1^2 + x_2^2\,. \tag{4.11}$$

Note that from Eq. (4.6) 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

$$\frac{{\rm d}x_1}{{\rm d}N} = -3x_1 + \frac{\sqrt{6}}{2}\lambda x_2^2 - x_1\frac{1}{H}\frac{{\rm d}H}{{\rm d}N} - \frac{\sqrt{6}}{2}Q\left(1 - x_1^2 - x_2^2 - x_3^2\right)\,, \tag{4.12}$$
$$\frac{{\rm d}x_2}{{\rm d}N} = -\frac{\sqrt{6}}{2}\lambda x_1 x_2 - x_2\frac{1}{H}\frac{{\rm d}H}{{\rm d}N}\,, \tag{4.13}$$
$$\frac{{\rm d}x_3}{{\rm d}N} = -2x_3 - x_3\frac{1}{H}\frac{{\rm d}H}{{\rm d}N}\,. \tag{4.14}$$

There are eight fixed points in total, see Table 4.1. 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 the problem [8.3]).

Among the eight fixed points presented in Table 4.1, we now identify the points responsible for radiation, matter, and accelerated eras.

[a] 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.

Name	x_1	x_2	x_3	Ω_ϕ	Ω_r	w_ϕ	$w_{\rm eff}$
(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)	$\frac{2\sqrt{6}}{3\lambda}$	$\frac{2\sqrt{3}}{3\lambda}$	$(1-\frac{4}{\lambda^2})^{1/2}$	$\frac{4}{\lambda^2}$	$1-\frac{4}{\lambda^2}$	$\frac{1}{3}$	$\frac{1}{3}$

Table 4.1: The fixed points for the coupled quintessence model with an exponential potential.

- (i) Radiation era

  The radiation-dominated epoch can be realized either by the points (e), (f), or (g) because they correspond to $w_{\rm eff}=1/3$. However the nucleosynthesis bound 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

  $$\mu=-1,\ 1,\ 2\,. \tag{4.15}$$

  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 (ϕMDE)" [17]. In order for the ϕMDE to be responsible for the matter era we require that $Q^2\ll1$ from the condition $\Omega_\phi=2Q^2/3\ll1$. The eigenvalues of the Jacobian matrix for perturbations about the point (a) is

  $$\mu=\frac{3}{2}+Q(Q+\lambda),\ -\frac{3}{2}+Q^2,\ -\frac{1}{2}+Q^2\,. \tag{4.16}$$

  As long as $Q^2\ll1$, two of the eigenvalues are negative. One of them is positive for $Q(Q+\lambda)>-3/2$, which is satisfied unless $Q<0$ and $\lambda\gg1$. Hence the ϕMDE is a saddle followed by a late-time accelerated point.

  Since the effective equation of state for the point (d) is given by $w_{\rm eff}=-Q/(Q+\lambda)$, it is possible to have $w_{\rm eff}\simeq0$ for $|\lambda|\gg|Q|$. The eigenvalues of the Jacobian matrix for perturbations about the point (d) are

  $$\mu=-\frac{4Q+\lambda}{2(Q+\lambda)},\ -\frac{3(2Q+\lambda)}{4(Q+\lambda)}\left[1\pm\sqrt{1+\frac{8[3-\lambda(Q+\lambda)][3+2Q(Q+\lambda)]}{3(2Q+\lambda)^2}}\right]. \tag{4.17}$$

  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

  $$\mu=\frac{1}{2}(\lambda^2-4),\ \frac{1}{2}(\lambda^2-6),\ \lambda(Q+\lambda)-3\,. \tag{4.18}$$

Under the condition $\lambda^2 < 2$, this point is stable for

$$\lambda(Q+\lambda) < 3\,. \tag{4.19}$$

The energy fraction of the field for the point (d) satisfies $\Omega_\phi > 1$ under the condition (4.19). For the point (d) the condition for the acceleration, $w_{\rm eff} < -1/3$, 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. (4.17) is larger than unity under the condition (4.19) with $-\frac{4Q+\lambda}{2(Q+\lambda)} < 0$ and $-\frac{3(2Q+\lambda)}{4(Q+\lambda)} < 0$. Hence one of the eigenvalues in Eq. (4.17) is positive, which means that the point (d) is a saddle if the point (c) is stable. If the condition

$$\lambda(Q+\lambda) > 3\,, \tag{4.20}$$

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 (4.20) 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 ϕMDE 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 ϕMDE 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 ϕMDE 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:

$$\text{(e)} \to \text{(a)} \to \text{(c)}\,. \tag{4.21}$$

The presence of the saddle ϕMDE demands the conditions $Q^2 \ll 1$ and $Q(Q+\lambda) > -3/2$. The stability of the accelerated points requires the conditions $\lambda^2 < 2$ and $\lambda(Q+\lambda) < 3$.

In Fig. 4.1 we plot the cosmological evolution of the density parameters $\Omega_\phi, \Omega_m, \Omega_r$ as well as the equations of state w_ϕ and $w_{\rm eff}$ for $\lambda = 0.1$ and $Q = 0.3$. This shows that the matter era is in fact replaced by the ϕMDE with $\Omega_\phi = w_{\rm eff} \simeq 2Q^2/3 \simeq 0.06$. The ϕMDE is followed by the accelerated point (c) with the future asymptotic values of the equations of state: $w_\phi = w_{\rm eff} = -1 + \lambda^2/3 \simeq -0.996$.

The presence of the ϕMDE changes the background expansion history of the universe. Since the evolution of the scale factor during the ϕMDE 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 is smaller than in the uncoupled case by roughly a factor $z_{\rm dec}^{Q^2}$. 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. A full comparison with 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].

4.2 Decoupling the baryons

The scalar field coupling induces a variation of the particle masses. As it can be seen from the conservation equation (4.4) the matter density varies as

$$\rho_m = \rho_m^{(0)}\,(a/a_0)^{-3} \exp\left(\int_{\phi_0}^{\phi} Q(\tilde{\phi})\mathrm{d}\tilde{\phi}\right)\,, \tag{4.22}$$

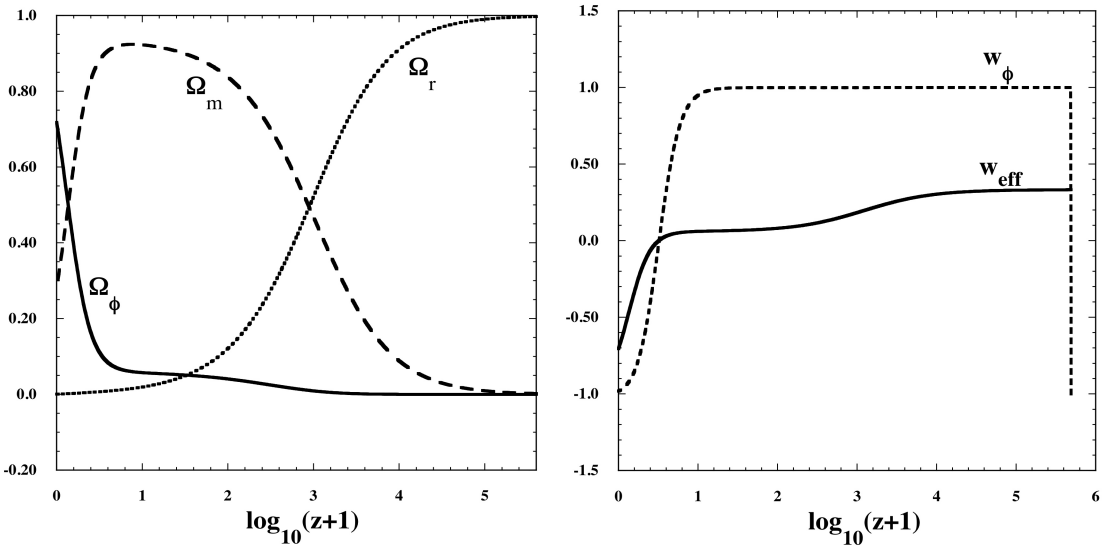

Figure 4.1: Cosmological evolution of the coupled quintessence scenario with an exponential potential for $\lambda = 0.1$ and $Q = 0.3$. The initial conditions are chosen to be $x_1 = 0$, $x_2 = 10^{-10}$, and $x_3 = 0.999$ at the redshift $\log_{10}(z+1) = 5.6919$. The field equation of state w_ϕ starts 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 ϕMDE (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 $w_{\rm eff} = w_\phi = -1 + \lambda^2/3$ and $\Omega_\phi = 1$.

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

$$m = m^{(0)} \exp\left(\int_{\phi_0}^{\phi} Q(\tilde{\phi}) \mathrm{d}\tilde{\phi} \right) . \tag{4.23}$$

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]

$$\left| \frac{\dot{G}}{G} \right| \lesssim \text{few} \times 10^{-11} \text{ yrs}^{-1} . \tag{4.24}$$

Then we have, in the case of a constant coupling Q,

$$\left| \frac{1}{G} \frac{\mathrm{d}G}{\mathrm{d}N} \right| = \left| \frac{1}{m} \frac{\mathrm{d}m}{\mathrm{d}N} \right| = \left| Q \frac{\mathrm{d}\phi}{\mathrm{d}N} \right| \lesssim 10^{-1} . \tag{4.25}$$

Note that we have used the present value of the Hubble parameter given in Eq. (1.42). For instance, on the solution (d) of the previous section, one has $\mathrm{d}\phi/\mathrm{d}N = \sqrt{6}x_1 = 3/(Q+\lambda)$ 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 acceleration ($w_{\rm eff} = -Q/(Q+\lambda) \approx 0$).

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. 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 (4.12)-(4.14) (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 4.1 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 solving 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 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 $\Omega_{\rm eff}$ with an effective constant equation of state $w_{\rm eff}$, i.e. $\rho_c \propto \rho_\phi \propto a^{-3(1+w_{\rm eff})}$. For the scaling solution (d) we have $w_{\rm eff} = -Q/(Q+\lambda)$. 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

$$a_b = \left(\Omega_b^{(0)}/\Omega_{\rm eff}^{(0)}\right)^{-1/(3w_{\rm eff})} , \tag{4.26}$$

and $\Omega_{\rm eff}^{(0)} = \Omega_c^{(0)} + \Omega_\phi^{(0)} = 1 - \Omega_b^{(0)}$ (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 density 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 $w_{\rm eff} \approx -1$, 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 $z_{\rm acc}$) by solving for $\ddot{a} = 0$ in the flat-space Friedmann equation:

$$\frac{\ddot{a}}{a} = -\frac{4\pi}{3}\left[\rho_{\rm eff}(1+3w_{\rm eff}) + \rho_b\right] = -\frac{1}{2}H_0^2\left[\Omega_{\rm eff}^{(0)} a^{-3(1+w_{\rm eff})}(1+3w_{\rm eff}) + (1-\Omega_{\rm eff}^{(0)})a^{-3}\right] = 0\,. \tag{4.27}$$

This amounts to

$$z_{\rm acc} = -1 + \left[\frac{\Omega_{\rm eff}^{(0)} - 1}{\Omega_{\rm eff}^{(0)}(1+3w_{\rm eff})}\right]^{1/(3w_{\rm eff})} , \tag{4.28}$$

which gives $z_{\rm acc} \approx 2.6$ for $\Omega_b^{(0)} = 0.04$ and $w_{\rm eff} = -1$. This redshift can increase up to $z_{\rm acc} \approx 4$ for $w_{\rm eff} = -0.6$, 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.

4.3 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:

$$\dot{\rho}_m + 3H\rho_m = +\Gamma\rho_m\,, \tag{4.29}$$

$$\dot{\rho}_X + 3H(1+w_X)\rho_X = -\Gamma\rho_m\,, \tag{4.30}$$

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

$$\delta \equiv \Gamma/H\,. \tag{4.31}$$

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

$$3H^2 = \rho_m + \rho_r + \rho_X \,. \tag{4.32}$$

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 (4.29) can be written in an integrated form

$$\rho_m = \rho_m^{(0)} (a/a_0)^{-3} \exp\left(\int \delta \, \mathrm{d}(\, \ln a) \right) . \tag{4.33}$$

The cosmological evolution is different depending on the form of the coupling δ. In the following we shall consider only the case of constant δ.

For constant δ, Eq. (4.33) is integrated to give

$$\rho_m = \rho_m^{(0)} (a/a_0)^{-3+\delta} = \rho_m^{(0)} (1+z)^{3-\delta} \,. \tag{4.34}$$

If w_X is constant, substituting Eq. (4.34) into (4.30) leads to the following equation

$$\rho_X = \rho_X^{(0)} (1+z)^{3(1+w_X)} + \rho_m^{(0)} \frac{\delta}{\delta + 3w_X} \left[(1+z)^{3(1+w_X)} - (1+z)^{3-\delta} \right] . \tag{4.35}$$

From the Friedmann equation (4.32) and neglecting radiation we obtain

$$E^2(z) = \Omega_X^{(0)} (1+z)^{3(1+w_X)} + \frac{1-\Omega_X^{(0)}}{\delta + 3w_X} \left[\delta (1+z)^{3(1+w_X)} + 3w_X (1+z)^{3-\delta} \right] , \tag{4.36}$$

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. (4.35) 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 (4.36) 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. 4.2 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 ΛCDM 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.

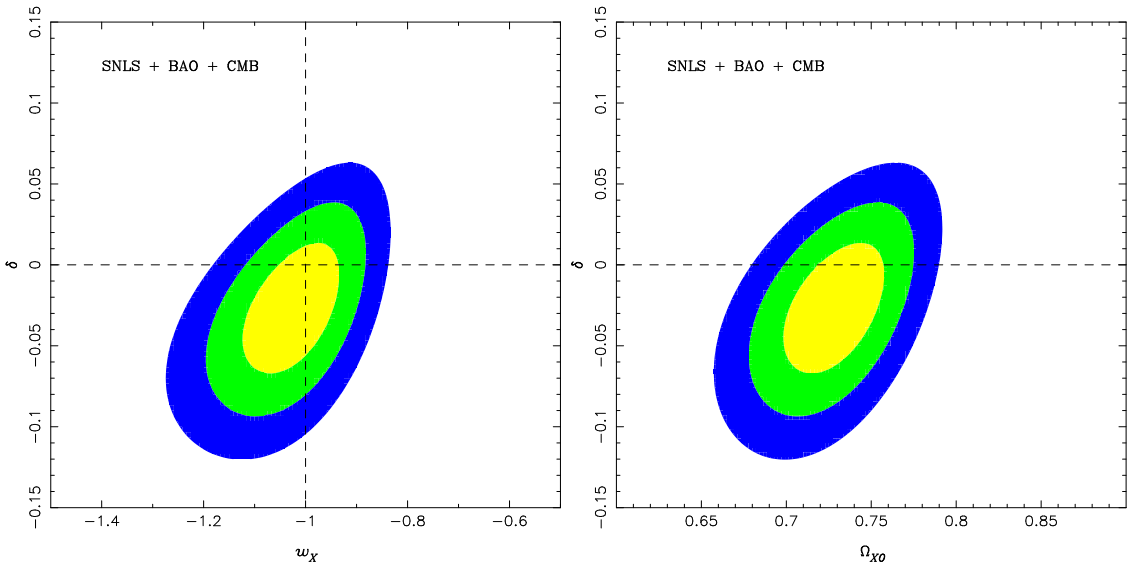

Figure 4.2: Probability contours (1σ, 2σ, and 3σ) for the constant δ parametrization (4.36) obtained from the combined data analysis of SNLS [107], the CMB shift parameter $\mathcal{R}$ [14], and the BAO effective distance D_V [68]. The left panel shows observational contours in the (w_X, δ) plane marginalized over $\Omega_X^{(0)}$, whereas the right panel depicts contours in the $(\Omega_X^{(0)}, \delta)$ plane marginalized over w_X. The ΛCDM 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].

Chapter 5

Modified gravity

Einstein's gravity is the most general theory of gravity in four dimensions with second-order equations of motion and a tensor field only. It can be generalized in many ways. 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:

$$S = \frac{1}{2\kappa^2}\int \mathrm{d}^4x\sqrt{-g}f(R) + S_m(g_{\mu\nu}, \Psi_m)\,, \tag{5.1}$$

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. Before we proceed we need however a mathematical tool.

5.1 Conformal transformations

We will use more than once in the following a simple mechanism to introduce a new degree of freedom in the Einstein Lagrangian, making use of a conformal transformation:

$$\tilde{g}_{\mu\nu} = f(\phi)g_{\mu\nu} \tag{5.2}$$

where ϕ is a scalar field and $f(\phi)$ a generic invertible (so $\tilde{g}^{\mu\nu}$ exists) and positive-definite function. The transformation is called conformal because angles between four-vectors remain the same in the two metrics. Notice that we are not transforming the coordinates, but the metric elements. Under this transformation, one has (all tilded quantities make use of $\tilde{g}_{\mu\nu}$)

$$\sqrt{-\tilde{g}} = f^2\sqrt{-g} \tag{5.3}$$

$$\tilde{R} = f^{-1}(R - 6\omega_{,\mu}g^{\mu\nu}\omega_{,\nu} - 6\Box\omega) \tag{5.4}$$

$$\tilde{\Box} = f^{-1}[2\omega_{,\mu}g^{\mu\nu}\partial_\nu + \Box] \tag{5.5}$$

with $\omega_{,\mu} \equiv f_{,\mu}/2f = f'\phi_\mu/(2f)$. Using these relations, one sees that for a model with a scalar field and ordinary matter (we assume here units such that the factor $1/16\pi G$ in front of R is unity) the Lagrangian can be transformed as follows

$$S = \int \mathrm{d}^4x\sqrt{-g}[R + L_\phi(g_{\mu\nu}) + L_m(g_{\mu\nu}, \Psi)] \tag{5.6}$$

$$= \int \mathrm{d}^4x\sqrt{-\tilde{g}}f^{-2}[f(\tilde{R} + 6\tilde{\Box}\omega - 6\tilde{g}^{\mu\nu}\omega_{,\mu}\omega_{,\nu}) + L_\phi(f^{-1}\tilde{g}_{\mu\nu}) + L_m(f^{-1}\tilde{g}_{\mu\nu}, \Psi)] \tag{5.7}$$

$$= \int \mathrm{d}^4x\sqrt{-\tilde{g}}\{f^{-1}\tilde{R} + [f^{-2}L_\phi(f^{-1}\tilde{g}_{\mu\nu}) + f^{-1}(6\tilde{\Box}\omega - 6\tilde{g}^{\mu\nu}\omega_{,\mu}\omega_{,\nu})] + f^{-2}L_m(f^{-1}\tilde{g}_{\mu\nu}, \Psi)\} \tag{5.8}$$

$$= \int \mathrm{d}^4x\sqrt{-\tilde{g}}\{f^{-1}\tilde{R} + \tilde{L}_\phi(\tilde{g}_{\mu\nu}) + \tilde{L}_m(f^{-1}\tilde{g}_{\mu\nu}, \Psi)\} = S' \tag{5.9}$$

where Ψ represents matter fields (baryons and DM) and $\tilde{L}_m = f^{-2}L_M$, and the new scalar Lagrangian is

$$\tilde{L}_\phi(\tilde{g}_{\mu\nu}) = f^{-2}L_\phi(f^{-1}\tilde{g}_{\mu\nu}) + f^{-1}(6\tilde{\Box}\omega - 6\tilde{g}^{\mu\nu}\omega_{,\mu}\omega_{,\nu}) \tag{5.10}$$

Note that even if one had $L_\phi = 0$ initially, the transformation would still induce a kinetic term for ϕ. If

$$L_\phi = -\frac{1}{2}\phi_{,\mu}g^{\mu\nu}\phi_{,\nu} - V(\phi) \tag{5.11}$$

then (using integration by parts)

$$\tilde{L}_\phi(\tilde{g}_{\mu\nu}) = f^{-2}L_\phi(f^{-1}\tilde{g}_{\mu\nu}) + f^{-1}(6\tilde{\Box}\omega - 6\tilde{g}^{\mu\nu}\omega_{,\mu}\omega_{,\nu}) \tag{5.12}$$

$$= -\frac{1}{2}f^{-1}\tilde{g}^{\mu\nu}[1 - 3\left(\frac{f'}{f}\right)^2]\phi_{,\mu}\phi_{,\nu} - f^{-2}V(\phi) \tag{5.13}$$

which can be written in fully canonical form

$$\tilde{L} = -\frac{1}{2}\psi_{,\mu}\psi_{,\nu}\tilde{g}^{\mu\nu} - U(\psi) \tag{5.14}$$

by defining a new field and a new potential

$$\psi_{,\mu} = f^{-1/2}[1 - 3\left(\frac{f'}{f}\right)^2]^{1/2}\phi_{,\mu}$$

$$U(\psi) = \frac{V(\phi)}{f^2(\phi)}|_{\phi(\psi)}$$

The new Lagrangian contains now a direct coupling between ϕ and the new Ricci scalar, and also between ϕ and the matter fields Ψ inside L_m. The two actions S and S' obviously describe the same physics, but the first one is said to live in the Einstein frame (Einstein's equations being the standard ones) and the second in the so-called Jordan frame.

There are two main uses of the transformations: *a*) choose a $f(\phi)$ in such a way to simplify the equations of motion, and *b*) choose a $f(\phi)$ to generate forms of modified gravity. Let us consider the first application. Let us assume we have a Lagrangian with an explicit coupling (these models, called scalar-tensor or Brans-Dicke models, will be studied in Sec. 5.4)

$$A = \int \mathrm{d}^4x\sqrt{-g}\left[\frac{1}{2\kappa^2}F(\phi)R - \frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi) + L_M(g_{\mu\nu},\Psi)\right] \tag{5.15}$$

This is then a Jordan frame since the EMT of the matter fields Ψ obeys the standard conservation laws. To convert the theory into the Einstein frame, we notice that the action A has the same structure as S of Eq. (5.6) except for the factor F. Repeating the same steps, one sees that by choosing the confornal factor $f = F$, i.e. with the transformation

$$\tilde{g}_{\mu\nu} = F(\phi)g_{\mu\nu}, \tag{5.16}$$

one obtains

$$A = \int \mathrm{d}^4x\sqrt{-\tilde{g}}[\tilde{R} + \tilde{L}_\phi(\tilde{g}_{\mu\nu},\phi) + \tilde{L}_m(F^{-1}\tilde{g}_{\mu\nu},\Psi)] \tag{5.17}$$

with suitable redefinitions of $\tilde{L}_\phi, \tilde{L}_m$. Now we are in an Einstein frame, where the gravitational sector is the Einstein-Hilbert Lagrangian, but the matter sector includes a coupling between ϕ and Ψ, and therefore matter particles will obey modified conservation equations. In fact, we know already that one obtains a form of coupled dark energy. Scalar-tensor and coupled dark energy models can indeed be transformed one into the other via conformal transformations.

However, one must keep in mind that when the equations of motion for the matter particles deviate from the standard ones, matter will appear to be non-conserved (eg. particle masses will change in time and space

or, more in general, the particles do not follow geodesics) and all the experimental/observational results will be more difficult to interpret. For instance, if the Rydberg constant

$$R = \frac{m_e e^4}{8\epsilon_0^2 h^3 c} \tag{5.18}$$

on which the measurement of redshifts depend, varied in space and time with the electron mass, then all estimation of redshifts would be problematic. Therefore usually it is always preferred to stay in that frame in which the matter EMT is conserved. If baryons and DM particles have different couplings, then there can be several possible frames, eg frames in which baryons are conserved but DM is not. In this case, the "observational" frame (sometimes called "physical" frame), i.e. the one in which experiments can immediately be compared to theory, is the one in which baryonic particles are conserved (and therefore the masses of eg protons, electrons etc. do not vary with time and space). It can also happen that although the Lagrangian seems to violate the standard conservation equations, these are recovered at the local level, i.e. around non-linear structures like the Earth, Sun, stars or the Milky Way halo, which is where indeed experiments are performed. This effect is called screened and will be studied later on.

As for the use of the metric transformations to generate new classes of theories, we can simply consider that if we start from

$$S = \int \mathrm{d}^4x\sqrt{-g}[R + L_m(g_{\mu\nu}, \Psi)] \tag{5.19}$$

and rewrite the action in the same way but using the transformed metric $\tilde{g}$, i.e. as

$$\tilde{S} = \int \mathrm{d}^4x\sqrt{-\tilde{g}}[\tilde{R} + L_m(\tilde{g}_{\mu\nu}, \Psi)] \tag{5.20}$$

where $\tilde{g} = \tilde{g}(g, \phi)$, we have created a different model (i.e. S and $\tilde{S}$ are not just two ways of writing the same action, as for S, S' in our initial example). In the new action ϕ is in fact an additional degree of freedom, and its variation will give rise to new equations of motion. This kind of coupling is called *conformal coupling*.

This can be immediately generalized: in fact, the transformation does not have to be a conformal one. For instance, the new metric $\tilde{g}$ might include derivatives of the scalar field, e.g.

$$\tilde{g}_{\mu\nu} = \phi_{,\mu}\phi_{,\nu} + f(\phi)g_{\mu\nu} \tag{5.21}$$

If both metrics are invertible, the procedure is acceptable and new classes of modified gravity can be generated. In fact, most theories of modified gravity currently investigated can be produced this way.

5.2 Dynamics of $f(R)$ models

Let us now go back to Eq. (5.1). There are two approaches to derive field equations from the action (5.1).

- **(I) The metric formalism**

  The first approach is the so-called metric formalism in which the connections $\Gamma^{\alpha}_{\beta\gamma}$ are the usual connections defined in terms of the metric $g_{\mu\nu}$. The field equations can be obtained by varying the action (5.1) with respect to $g_{\mu\nu}$:

  $$F(R)R_{\mu\nu}(g) - \frac{1}{2}f(R)g_{\mu\nu} - \nabla_\mu\nabla_\nu F(R) + g_{\mu\nu}\Box F(R) = \kappa^2 T_{\mu\nu}\,, \tag{5.22}$$

  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 the ones employed to obtain the field equation (2.7). The trace of Eq. (5.22) is given by

  $$3\Box F(R) + F(R)R - 2f(R) = \kappa^2 T\,, \tag{5.23}$$

  where $T = g^{\mu\nu}T_{\mu\nu} = -\rho + 3P$. Here ρ and P are the energy density and the pressure of the matter, respectively.

- **(II) The Palatini formalism**

  The second approach is the so-called Palatini formalism in which $\Gamma^{\alpha}_{\beta\gamma}$ and $g_{\mu\nu}$ are treated as independent variables. Varying the action (5.1) with respect to $g_{\mu\nu}$ gives

$$F(R)R_{\mu\nu}(\Gamma) - \frac{1}{2}f(R)g_{\mu\nu} = \kappa^2 T_{\mu\nu}\,, \tag{5.24}$$

  where $R_{\mu\nu}(\Gamma)$ is the Ricci tensor corresponding to the connections $\Gamma^{\alpha}_{\beta\gamma}$. In general this is different from the Ricci tensor $R_{\mu\nu}(g)$ corresponding to the metric connections. Taking the trace of Eq. (5.24), we obtain

$$F(R)R - 2f(R) = \kappa^2 T\,, \tag{5.25}$$

  where $R(T) = g^{\mu\nu}R_{\mu\nu}(\Gamma)$ is directly related to T. Taking the variation of the action (5.1) with respect to the connection, and using Eq. (5.24), we find

$$\begin{aligned} R_{\mu\nu}(g) - \frac{1}{2}g_{\mu\nu}R(g) &= \frac{\kappa^2 T_{\mu\nu}}{F} - \frac{FR(T)-f}{2F}g_{\mu\nu} + \frac{1}{F}(\nabla_\mu\nabla_\nu F - g_{\mu\nu}\Box F) \\ &\quad -\frac{3}{2F^2}\left[\partial_\mu F\partial_\nu F - \frac{1}{2}g_{\mu\nu}(\nabla F)^2\right]. \end{aligned} \tag{5.26}$$

In General Relativity we have $f(R) = R - 2\Lambda$ and $F(R) = 1$, so that the term $\Box F(R)$ in Eq. (5.23) 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 $\Box F(R)$ does not vanish in Eq. (5.23). This means that, in the metric formalism, there is a propagating scalar degree of freedom, $\psi \equiv F(R)$. The trace equation (5.23) governs the dynamics of the scalar field ψ (dubbed "scalaron" [8]). In the Palatini formalism the kinetic term $\Box F(R)$ is not present in Eq. (5.25), which means that the scalar-field degree of freedom does not propagate freely.

The de Sitter point corresponds to a vacuum solution at which the Ricci scalar is constant. Since $\Box F(R) = 0$ at this point, we get

$$F(R)R - 2f(R) = 0\,, \tag{5.27}$$

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].

The dynamics of $f(R)$ dark energy models is different depending on the two formalisms, but we will confine ourselves to the metric case.

5.3 $f(R)$ gravity in 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, 354]. 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 $f_{,RR}$ [355, 356, 357, 358, 359]. 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 [356, 358, 359, 360], for the presence of the matter-dominated epoch [272, 361], and for the stability of cosmological perturbations [362, 363, 364, 365].

- (iii) $f(R) \to R - 2\Lambda$ for $R \gg R_0$.

  This is required for consistency with local gravity tests [266, 366, 367, 368, 369] and for the presence of the matter-dominated epoch [361].

- (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 [370, 371, 361].

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.

$$\text{(A) } f(R) = R - \mu R_c (R/R_c)^p \qquad \text{with } \ 0 < p < 1, \ \ \mu, R_c > 0\,, \tag{5.28}$$

$$\text{(B) } f(R) = R - \mu R_c \frac{(R/R_c)^{2n}}{(R/R_c)^{2n} + 1} \qquad \text{with } \ n, \mu, R_c > 0\,, \tag{5.29}$$

$$\text{(C) } f(R) = R - \mu R_c \left[1 - \left(1 + R^2/R_c^2\right)^{-n}\right] \qquad \text{with } \ n, \mu, R_c > 0\,, \tag{5.30}$$

$$\text{(D) } f(R) = R - \mu R_c \tanh\left(R/R_c\right) \qquad \text{with } \ \mu, R_c > 0\,. \tag{5.31}$$

The models (A), (B), (C), and (D) have been proposed in Refs. [361], [366], [367], and [369], respectively. A model similar to (D) has been also proposed in Ref. [368], while a generalized model encompassing (B) and (C) has been studied in Ref. [373]. 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 $f(R) \to R - \mu R_c[1 - (R^2/R_c^2)^{-n}]$ 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 $f(R) \to R - \mu R_c$ in the region $R \gg R_c$. These models satisfy $f(R = 0) = 0$, so the cosmological contant vanishes in the flat spacetime.

Let us consider now in more detail the cosmological dynamics of $f(R)$ gravity. 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

$$R = 6(2H^2 + \dot{H})\,, \tag{5.32}$$

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. (5.22) and (5.23) we obtain the following equations

$$3FH^2 = \kappa^2(\rho_m + \rho_r) + (FR - f)/2 - 3H\dot{F}\,, \tag{5.33}$$

$$-2F\dot{H} = \kappa^2[\rho_m + (4/3)\rho_r] + \ddot{F} - H\dot{F}\,. \tag{5.34}$$

We introduce the dimensionless variables:

$$x_1 \equiv -\frac{\dot{F}}{HF}\,, \quad x_2 \equiv -\frac{f}{6FH^2}\,, \quad x_3 \equiv \frac{R}{6H^2}\,, \quad x_4 \equiv \frac{\kappa^2\rho_r}{3FH^2}\,, \tag{5.35}$$

together with the following quantities

$$\Omega_m \equiv \frac{\kappa^2\rho_m}{3FH^2} = 1 - x_1 - x_2 - x_3 - x_4\,, \qquad \Omega_r \equiv x_4\,, \qquad \Omega_{\rm DE} \equiv x_1 + x_2 + x_3\,. \tag{5.36}$$

It is straightforward to derive the following differential equations [361]:

$$\frac{\mathrm{d}x_1}{\mathrm{d}N} = -1 - x_3 - 3x_2 + x_1^2 - x_1 x_3 + x_4 \,, \tag{5.37}$$

$$\frac{\mathrm{d}x_2}{\mathrm{d}N} = \frac{x_1 x_3}{m} - x_2(2x_3 - 4 - x_1) \,, \tag{5.38}$$

$$\frac{\mathrm{d}x_3}{\mathrm{d}N} = -\frac{x_1 x_3}{m} - 2x_3(x_3 - 2) \,, \tag{5.39}$$

$$\frac{\mathrm{d}x_4}{\mathrm{d}N} = -2x_3 x_4 + x_1 x_4 \,, \tag{5.40}$$

where $N = \ln a$ and

$$m \equiv \frac{\mathrm{d}\,\ln F}{\mathrm{d}\,\ln R} = \frac{R f_{,RR}}{f_{,R}} \,, \tag{5.41}$$

$$r \equiv -\frac{\mathrm{d}\,\ln f}{\mathrm{d}\,\ln R} = -\frac{R f_{,R}}{f} = \frac{x_3}{x_2} \,. \tag{5.42}$$

From Eq. (5.42) 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 ΛCDM model, $f(R) = R - 2\Lambda$, corresponds to $m = 0$. Hence the quantity m characterizes the deviation from the ΛCDM model. Note also that the model, $f(R) = \alpha R^{1+m} - 2\Lambda$, gives a constant value of m. The analysis using Eqs. (5.37)-(5.40) 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 is

$$w_{\rm eff} = -\frac{1}{3}(2x_3 - 1) \,, \tag{5.43}$$

In the absence of radiation ($x_4 = 0$) the fixed points for the dynamical system (5.37)-(5.40) are

$$P_1 : (x_1, x_2, x_3) = (0, -1, 2), \quad \Omega_m = 0, \quad w_{\rm eff} = -1 \,, \tag{5.44}$$

$$P_2 : (x_1, x_2, x_3) = (-1, 0, 0), \quad \Omega_m = 2, \quad w_{\rm eff} = 1/3 \,, \tag{5.45}$$

$$P_3 : (x_1, x_2, x_3) = (1, 0, 0), \quad \Omega_m = 0, \quad w_{\rm eff} = 1/3 \,, \tag{5.46}$$

$$P_4 : (x_1, x_2, x_3) = (-4, 5, 0), \quad \Omega_m = 0, \quad w_{\rm eff} = 1/3 \,, \tag{5.47}$$

$$P_5 : (x_1, x_2, x_3) = \left(\frac{3m}{1+m}, -\frac{1+4m}{2(1+m)^2}, \frac{1+4m}{2(1+m)} \right) ,$$
$$\Omega_m = 1 - \frac{m(7+10m)}{2(1+m)^2}, \quad w_{\rm eff} = -\frac{m}{1+m}, \tag{5.48}$$

$$P_6 : (x_1, x_2, x_3) = \left(\frac{2(1-m)}{1+2m}, \frac{1-4m}{m(1+2m)}, -\frac{(1-4m)(1+m)}{m(1+2m)} \right) ,$$
$$\Omega_m = 0, \quad w_{\rm eff} = \frac{2 - 5m - 6m^2}{3m(1+2m)}. \tag{5.49}$$

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

$$\{x_1, x_2, x_3\} = \left\{ \frac{3m(x_3/x_2)}{1 + m(x_3/x_2)}, -\frac{1 + 4m(x_3/x_2)}{2[1 + m(x_3/x_2)]^2}, \frac{1 + 4m(x_3/x_2)}{2[1 + m(x_3/x_2)]} \right\} , \tag{5.50}$$

for P_5 and an analogous set for P_6. We will call $m_{5,6}$ the solutions of these equations and

$$m'_{5,6} \equiv \left. \frac{\mathrm{d}m}{\mathrm{d}r} \right|_{m_{5,6}} , \tag{5.51}$$

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 $w_{\rm eff} \simeq 0$, 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 ϕMDE [17], but in this case the ϕMDE cannot be responsible for the matter fixed point because $w_{\rm eff}(P_2) = 1/3$. 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 [361]. For the point P_5 the eigenvalues for the 3×3 Jacobian matrix of perturbations are

$$3(1+m_5'), \quad \frac{-3m_5 \pm \sqrt{m_5(256m_5^3+160m_5^2-31m_5-16)}}{4m_5(m_5+1)}\,, \tag{5.52}$$

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 $m_5 \to -0$, 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. (5.52) 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

$$m(r) \approx +0\,, \quad \frac{{\rm d}m}{{\rm d}r} > -1\,, \quad \text{at} \quad r = -1\,. \tag{5.53}$$

The first condition implies that the $f(R)$ models need to be close to the ΛCDM 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 (5.27) is satisfied in this case. The eigenvalues for the 3×3 matrix of perturbations about the point P_1 are

$$-3, \quad -\frac{3}{2} \pm \frac{\sqrt{25-16/m_1}}{2}\,, \tag{5.54}$$

where $m_1 = m(r = -2)$. This shows that the condition for the stability of the de Sitter point P_1 is given by

$$0 < m(r = -2) < 1\,, \tag{5.55}$$

which corresponds to the condition (iv) given in Sec. 5.3. The trajectories which start from the saddle matter point P_5 [satisfying the condition (5.53)] and then approach the stable de Sitter point P_1 [satisfying the condition (5.55)] are cosmologically viable.

The point P_6 is on the line $m(r) = -r - 1$. It can satisfy the condition for the cosmic acceleration ($w_{\rm eff} < -1/3$) 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

$$-4 + \frac{1}{m_6}, \quad \frac{2-3m_6-8m_6^2}{m_6(1+2m_6)}, \quad -\frac{2(m_6^2-1)(1+m_6')}{m_6(1+2m_6)}\,. \tag{5.56}$$

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, $w_{\rm eff} > -1$. One has $w_{\rm eff} \to -1$ in the limit $m_6 \to -\infty$.
- (b) $-1/2 < m_6 < 0$: P_6 corresponds to a strongly phantom behavior with $w_{\rm eff} < -7.6$.
- (c) $m_6 \geq 1$: P_6 corresponds to a slightly phantom behavior with $-1.07 < w_{\rm eff} \leq -1$. One has $w_{\rm eff} \to -1$ in the limit $m_6 \to +\infty$ and $m_6 \to 1$.

[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 $w_{\rm eff} > -1$.

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. 5.1 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 ΛCDM 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. 5.1 represents the following model [266]

$$f(R) = (R^b - \Lambda)^c \,, \tag{5.57}$$

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 $f(R) = R - \mu R_c (R/R_c)^p$ $(0 < p < 1, \mu, R_c > 0)$ [361, 374], which corresponds to the curve $m = p(1 + r)/r$. These models fall into Class A.

The models (5.29) and (5.30) 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

$$m(r) = C(-r-1)^{2n+1} \,, \tag{5.58}$$

where $C = 2n(2n+1)/\mu^{2n}$. The parameter $m(r)$ rapidly approaches 0 in the limit $r \to -1$ 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. 5.1 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. 5.1 are a couple of representative models giving viable cosmological evolution.

The requirement $m \to +0$ during the radiation and matter eras means that the models need to be close to the ΛCDM 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 Sec. 5.3. 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 Sec. 5.3, which is also required for the consistency with local gravity constraints (as we will see later).

For the model (5.29) 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:

$$\mu = \frac{(1 + x_1^{2n})^2}{x_1^{2n-1}(2 + 2x_1^{2n} - 2n)} \,, \tag{5.59}$$

where $x_1 \equiv R_1/R_c$. From the stability condition $0 < m(r = -2) < 1$ we obtain

$$2x_1^{4n} - (2n-1)(2n+4)x_1^{2n} + (2n-1)(2n-2) > 0 \,. \tag{5.60}$$

When $n = 1$, for example, we have $x_1 > \sqrt{3}$ and $\mu > 8\sqrt{3}/9$. Under Eq. (5.60) 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. (5.59) 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 (5.29) is close to the ΛCDM model with the asymptotic form $f(R) \simeq R - \mu R_c \left[1 - (R/R_c)^{-2n}\right]$. The deviation from the ΛCDM model becomes important when R decreases to the order of R_c. Note that the model (5.30) also has a similar property.

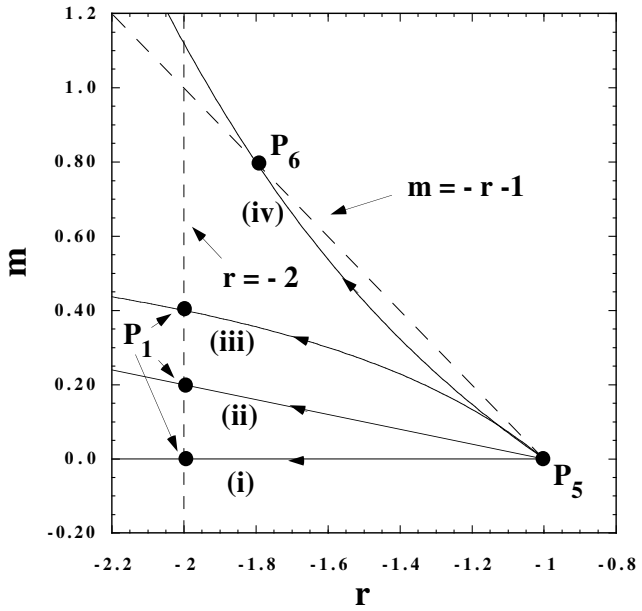

Figure 5.1: Four trajectories in the (r, m) plane. Each trajectory corresponds to the models: (i) ΛCDM, (ii) $f(R) = (R^b - \Lambda)^c$, (iii) $f(R) = R - \mu R_c (R/R_c)^p$ $(0 < p < 1, \mu, R_c > 0)$, and (iv) $m(r) = -C(r+1)(r^2 + ar + b)$.

In order to derive the equation of state of dark energy to confront with SN Ia observations, we rewrite Eqs. (5.33) and (5.34) as follows:

$$3AH^2 = \kappa^2 \left(\rho_m + \rho_r + \rho_{\rm DE}\right) , \tag{5.61}$$

$$-2A\dot{H} = \kappa^2 \left[\rho_m + (4/3)\rho_r + \rho_{\rm DE} + P_{\rm DE}\right] , \tag{5.62}$$

where A is some constant and

$$\kappa^2 \rho_{\rm DE} \equiv (1/2)(FR - f) - 3H\dot{F} + 3H^2(A - F) , \tag{5.63}$$

$$\kappa^2 P_{\rm DE} \equiv \ddot{F} + 2H\dot{F} - (1/2)(FR - f) - (3H^2 + 2\dot{H})(A - F) . \tag{5.64}$$

Defining $\rho_{\rm DE}$ and $P_{\rm DE}$ in the above way, one can show that these satisfy the usual conservation equations

$$\dot{\rho}_{\rm DE} + 3H(\rho_{\rm DE} + P_{\rm DE}) = 0 . \tag{5.65}$$

A similar procedure can be carried out for a more general Lagrangian density, $f(R, \phi, X)$. The dark energy equation of state, $w_{\rm DE} \equiv P_{\rm DE}/\rho_{\rm DE}$, is directly related to the one used in the standard analysis of SN Ia observations. From Eqs. (5.61) and (5.62) it is given by

$$w_{\rm DE} = -\frac{2A\dot{H} + 3AH^2 + \kappa^2 \rho_r/3}{3AH^2 - \kappa^2(\rho_m + \rho_r)} \simeq \frac{w_{\rm eff}}{1 - \tilde{\Omega}_m} , \tag{5.66}$$

where

$$\tilde{\Omega}_m \equiv \frac{\kappa^2 \rho_m}{3AH^2} = \frac{F}{A}\Omega_m . \tag{5.67}$$

The last approximate equality in Eq. (5.66) 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 ΛCDM model in the past, i.e. $F \to 1$ as $R \to \infty$. In order to reproduce the standard matter era for $z \gg 1$, we can choose $A = 1$ in Eqs. (5.61) and (5.62). 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 [372]). In both cases the equation of state changes from $w_{\rm DE} < -1$ to $w_{\rm DE} > -1$ before reaching the de Sitter attractor for viable $f(R)$ models [366, 266, 369]. 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. The models (5.29)-(5.31) are carefully constructed to have a suppressed m in the early cosmological epoch, while an appreciable deviation from the ΛCDM model ($m \lesssim \mathcal{O}(0.1)$) can appear around the present. Note that the model (5.28) does not allow this rapid evolution of m.

5.4 Scalar-tensor theories

As we have seen, 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 [398]). 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 $(\hbar^2 H_0 c^5/G)^{1/3} \simeq 68\, h^{1/3}\,\mathrm{MeV}$ or $(e^4 H_0 c^3/G)^{1/3} \simeq 2.5\, h^{1/3}\,\mathrm{MeV}$ are of the order of typical particle masses.

Brans-Dicke theory is just a particular example of scalar-tensor theories. These are relatively simple 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, 372, 399, 400] 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

$$S = \int \mathrm{d}^4x\sqrt{-g}\left[\frac{1}{2}f(\varphi, R) - \frac{1}{2}\zeta(\varphi)(\nabla\varphi)^2\right] + S_m(g_{\mu\nu}, \Psi_m)\,, \tag{5.68}$$

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 (5.68) 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 $\zeta = \omega_{\rm BD}/\varphi$, where $\omega_{\rm BD}$ 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 $\zeta = \omega_{\rm BD}/\varphi$. The action (5.68) can be transformed to the Einstein frame under the conformal transformation $\tilde{g}_{\mu\nu} = F g_{\mu\nu}$ with the choice

$$F \equiv \frac{\partial f}{\partial R}\,, \tag{5.69}$$

where F is positive in order to ensure that gravity is attractive.

Let us now consider theories of the type

$$f(\varphi, R) = F(\varphi)R - 2U(\varphi)\,, \tag{5.70}$$

for which the conformal factor depends on φ only. Following the transformation, we obtain the action in the Einstein frame:

$$S_E = \int \mathrm{d}^4x\sqrt{-\tilde{g}}\left[\frac{1}{2}\tilde{R} - \frac{1}{2}(\tilde{\nabla}\phi)^2 - V(\phi)\right] + S_m(\tilde{g}_{\mu\nu}F^{-1}, \Psi_m)\,, \tag{5.71}$$

where we have introduced a new scalar field ϕ in order to make the kinetic term canonical:

$$\phi = \int \mathrm{d}\varphi \sqrt{\frac{3}{2}\left(\frac{F_{,\varphi}}{F}\right)^2 + \frac{\zeta}{F}}\,. \tag{5.72}$$

The potential $V(\phi)$ is given by

$$V = U/F^2\,. \tag{5.73}$$

In order to describe the strength of the coupling between dark energy and non-relativistic matter, we introduce the following quantity

$$Q \equiv -\frac{F_{,\phi}}{2F} = -\frac{F_{,\varphi}}{2F}\left[\frac{3}{2}\left(\frac{F_{,\varphi}}{F}\right)^2 + \frac{\zeta}{F}\right]^{-1/2} . \tag{5.74}$$

If Q is a constant, the following relations hold from Eqs. (5.72) and (5.74):

$$F = e^{-2Q\phi}, \quad \zeta = (1-6Q^2)F\left(\frac{\mathrm{d}\phi}{\mathrm{d}\varphi}\right)^2 . \tag{5.75}$$

Then the action (5.68) in the Jordan frame yields [273]

$$S = \int \mathrm{d}^4x\sqrt{-g}\left[\frac{1}{2}F(\phi)R - \frac{1}{2}(1-6Q^2)F(\phi)(\nabla\phi)^2 - U(\phi)\right] + S_m(g_{\mu\nu}, \Psi_m) . \tag{5.76}$$

In the limit $Q \to 0$ the action (5.76) reduces to the one for a minimally coupled scalar field ϕ with the potential $U(\phi)$. The transformation of the Jordan frame action (5.76) via a conformal transformation $\tilde{g}_{\mu\nu} = F(\phi)g_{\mu\nu}$ gives rise to the Einstein frame action (5.71) with a constant coupling Q.

It is instructive to compare (5.76) with the action of Brans-Dicke theory with a potential U:

$$S = \int \mathrm{d}^4x\sqrt{-g}\left[\frac{1}{2}\psi R - \frac{\omega_{\rm BD}}{2\psi}(\nabla\psi)^2 - U(\psi)\right] + S_m(g_{\mu\nu}, \Psi_m) . \tag{5.77}$$

Setting $\psi = F = e^{-2Q\phi}$, one easily finds that the two actions are equivalent if the parameter $\omega_{\rm BD}$ is related to Q via the relation [273]

$$3 + 2\omega_{\rm BD} = \frac{1}{2Q^2} . \tag{5.78}$$

Under this condition, the theories given by (5.76) are equivalent to Brans-Dicke theory with a potential U. In the General Relativistic limit, $Q \to 0$, we have $\omega_{\rm BD} \to \infty$ as expected.

Taking the variation of Eq. (5.77) with respect to $g_{\mu\nu}$ and ψ leads to the following equations

$$\begin{aligned} R_{\mu\nu}(g) - \frac{1}{2}g_{\mu\nu}R(g) &= \frac{1}{\psi}T_{\mu\nu} - \frac{1}{\psi}g_{\mu\nu}U(\psi) + \frac{1}{\psi}(\nabla_\mu\nabla_\nu\psi - g_{\mu\nu}\Box\psi) \\ &\quad + \frac{\omega_{\rm BD}}{\psi^2}\left[\partial_\mu\psi\partial_\nu\psi - \frac{1}{2}g_{\mu\nu}(\nabla\psi)^2\right] , \end{aligned} \tag{5.79}$$

$$(3+2\omega_{\rm BD})\Box\psi + 4U(\psi) - 2\psi U_{,\psi} = T . \tag{5.80}$$

The above discussion shows that the action (5.76) 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 nonminimally 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:

$$Q(\varphi) = \frac{\xi\varphi}{[1 - \xi\varphi^2(1-6\xi)]^{1/2}} . \tag{5.81}$$

Note that $Q \simeq \xi\varphi$ for $|\xi| \ll 1$ and $Q \simeq \pm 1/\sqrt{6}$ in the limit $|\xi| \gg 1$.

We shall now study the cosmological dynamics for the Jordan frame action (5.76) 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 (5.76) with respect to $g_{\mu\nu}$ and ϕ leads to the following equations of motion

$$3FH^2 = \tfrac{1}{2}(1-6Q^2)F\dot{\phi}^2 + U - 3H\dot{F} + \rho_m + \rho_r \,, \tag{5.82}$$

$$2F\dot{H} = -(1-6Q^2)F\dot{\phi}^2 - \ddot{F} + H\dot{F} - \rho_m - (4/3)\rho_r \,, \tag{5.83}$$

$$(1-6Q^2)F\left[\ddot{\phi} + 3H\dot{\phi} + (\dot{F}/2F)\dot{\phi}\right] + U_{,\phi} + QFR = 0 \,. \tag{5.84}$$

We introduce the following variables

$$x_1 \equiv \frac{\dot{\phi}}{\sqrt{6}H} \,, \quad x_2 \equiv \frac{1}{H}\sqrt{\frac{U}{3F}} \,, \quad x_3 \equiv \frac{1}{H}\sqrt{\frac{\rho_r}{3F}} \,, \tag{5.85}$$

and

$$\Omega_m \equiv \frac{\rho_m}{3FH^2} \,, \qquad \Omega_r \equiv x_3^2 \,, \qquad \Omega_{\rm DE} \equiv (1-6Q^2)x_1^2 + x_2^2 + 2\sqrt{6}Qx_1 \,. \tag{5.86}$$

These satisfy the relation $\Omega_m + \Omega_r + \Omega_{\rm DE} = 1$ from Eq. (5.82).

Using Eqs. (5.82)-(5.84), we obtain the differential equations for x_1, x_2 and x_3:

$$\frac{{\rm d}x_1}{{\rm d}N} = \frac{\sqrt{6}}{2}(\lambda x_2^2 - \sqrt{6}x_1) + \frac{\sqrt{6}Q}{2}\left[(5-6Q^2)x_1^2 + 2\sqrt{6}Qx_1 - 3x_2^2 + x_3^2 - 1\right] - x_1\frac{\dot{H}}{H^2} \,, \tag{5.87}$$

$$\frac{{\rm d}x_2}{{\rm d}N} = \frac{\sqrt{6}}{2}(2Q-\lambda)x_1x_2 - x_2\frac{\dot{H}}{H^2} \,, \tag{5.88}$$

$$\frac{{\rm d}x_3}{{\rm d}N} = \sqrt{6}Qx_1x_3 - 2x_3 - x_3\frac{\dot{H}}{H^2} \,, \tag{5.89}$$

where

$$\lambda \equiv -U_{,\phi}/U \,, \tag{5.90}$$

and $\dot{H}/H^2$ is given by

$$\frac{\dot{H}}{H^2} = -\frac{1-6Q^2}{2}\left(3 + 3x_1^2 - 3x_2^2 + x_3^2 - 6Q^2x_1^2 + 2\sqrt{6}Qx_1\right) + 3Q(\lambda x_2^2 - 4Q) \,. \tag{5.91}$$

The effective equation of state of the system is

$$w_{\rm eff} = -1 + \frac{1-6Q^2}{3}(3 + 3x_1^2 - 3x_2^2 + x_3^2 - 6Q^2x_1^2 + 2\sqrt{6}Qx_1) - 2Q(\lambda x_2^2 - 4Q) \,. \tag{5.92}$$

We now assume $\lambda = const$, i.e. an exponential potential. In the absence of radiation ($x_3 = 0$), the fixed points of the system (5.87)-(5.89) for constant λ are given by [273]

- (a) ϕMDE

$$(x_1, x_2) = \left(\frac{\sqrt{6}Q}{3(2Q^2-1)}, 0\right) \,, \quad \Omega_m = \frac{3-2Q^2}{3(1-2Q^2)^2} \,, \quad w_{\rm eff} = \frac{4Q^2}{3(1-2Q^2)} \,. \tag{5.93}$$

- (b) Kinetic points

$$(x_1, x_2) = \left(\frac{1}{\sqrt{6}Q \pm 1}, 0\right) \,, \quad \Omega_m = 0 \,, \quad w_{\rm eff} = \frac{3 \mp \sqrt{6}Q}{3(1 \pm \sqrt{6}Q)} \,. \tag{5.94}$$

- (c) Scalar-field dominated point

$$(x_1, x_2) = \left(\frac{\sqrt{6}(4Q-\lambda)}{6(4Q^2 - Q\lambda - 1)}, \left[\frac{6 - \lambda^2 + 8Q\lambda - 16Q^2}{6(4Q^2 - Q\lambda - 1)^2}\right]^{1/2}\right) \,, \quad \Omega_m = 0 \,,$$

$$w_{\rm eff} = -\frac{20Q^2 - 9Q\lambda - 3 + \lambda^2}{3(4Q^2 - Q\lambda - 1)} \,. \tag{5.95}$$

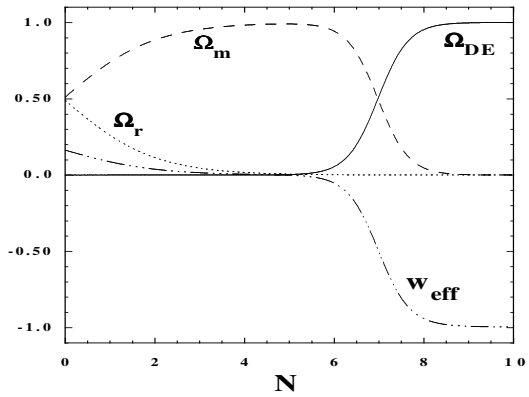

Figure 5.2: The evolution of $\Omega_{\rm DE}$, Ω_m, Ω_r, and $w_{\rm eff}$ with parameters $Q = 0.01$, $p = 0.2$, and $C = 0.7$. The initial conditions are given by $x_1 = 0$, $x_2 = 2.27 \times 10^{-7}$, $x_3 = 0.7$, and $x_4 - 1 = -5.0 \times 10^{-13}$ at $N = 0$. From Ref. [273].

- (d) Scaling solution

$$(x_1, x_2) = \left(\frac{\sqrt{6}}{2\lambda}, \sqrt{\frac{3 + 2Q\lambda - 6Q^2}{2\lambda^2}} \right), \quad \Omega_m = 1 - \frac{3 - 12Q^2 + 7Q\lambda}{\lambda^2}, \quad w_{\rm eff} = -\frac{2Q}{\lambda}. \tag{5.96}$$

- (e) de Sitter point (present for $\lambda = 4Q$)

$$(x_1, x_2) = (0, 1)\,, \quad \Omega_m = 0\,, \quad w_{\rm eff} = -1\,. \tag{5.97}$$

One can confirm that the de Sitter point (e) exists for $\lambda = 4Q$, by setting $\dot{\phi} = 0$ in Eqs. (5.82)-(5.84). This is the special case of the scalar-field dominated point (c).

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. 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 $w_{\rm eff} \simeq 4Q^2/3$. 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, $w_{\rm eff} \simeq 0$ for $|Q| \ll |\lambda|$. In this case, however, the condition $w_{\rm eff} < -1/3$ 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 Sec. 4.

5.5 Relation between $f(R)$ and scalar-tensor theories

Models of $f(R)$ gravity can be recast as scalar-tensor models via the conformal transformation

$$\tilde{g}_{\mu\nu} = \Omega^2 g_{\mu\nu}\,, \tag{5.98}$$

The action (5.1) can in fact be written as

$$S = \int \mathrm{d}^4x\sqrt{-g}\left(\frac{1}{2\kappa^2}FR - U\right) + S_m(g_{\mu\nu}, \Psi_m)\,, \tag{5.99}$$

where

$$U = \frac{RF - f}{2\kappa^2}\,. \tag{5.100}$$

The action (5.99) is then transformed to be

$$S = \int \mathrm{d}^4x\sqrt{-\tilde{g}}\left[\frac{1}{2\kappa^2}F\Omega^{-2}(\tilde{R} + 6\tilde{\Box}\omega - 6\tilde{g}^{\mu\nu}\omega_{,\mu}\omega_{,\nu}) - \Omega^{-4}U\right] + S_m(g_{\mu\nu}, \Psi_m)\,. \tag{5.101}$$

So we see that with

$$\Omega^2 = F \tag{5.102}$$

we recover an Einstein Lagrangian. We also introduce a new scalar field ϕ defined by

$$\kappa\phi \equiv \sqrt{\frac{3}{2}}\ln F\,. \tag{5.103}$$

Since $\Omega = \sqrt{F}$ and $\omega_{,\mu} = \Omega_{,\mu}/\Omega$, it follows that $\omega_{,\mu} = (1/\sqrt{6})\kappa\phi_{,\mu}$. The integral $\int \mathrm{d}^4x\sqrt{-\tilde{g}}\,\tilde{\Box}\omega$ vanishes on account of Gauss's theorem. Then the action in the Einstein frame is

$$S_E = \int \mathrm{d}^4x\sqrt{-\tilde{g}}\left[\frac{1}{2\kappa^2}\tilde{R} - \frac{1}{2}\tilde{g}^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi)\right] + S_m(g_{\mu\nu}, \Psi_m)\,, \tag{5.104}$$

where

$$V(\phi) = \frac{RF - f}{2\kappa^2F^2}\,. \tag{5.105}$$

As in the previous section, it is useful to write the coupling in the form

$$F = e^{-2Q\phi}\,, \tag{5.106}$$

From Eq. (5.103) we then find that $f(R)$ models are equivalent to scalar-tensor models with the specific coupling

$$Q = -1/\sqrt{6}\,. \tag{5.107}$$

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. (5.78) that in this case the Brans-Dicke parameter ω_{BD} is equivalent to 0.

5.6 Horndeski Lagrangian [a]

The scalar-tensor Lagrangian can be further generalized. In fact, Horndeski in 1974 found the most general scalar-tensor theory that gives second order equations of motion, the so-called Horndeski Lagrangian (HL). We give here a very brief introduction.

The HL is defined as the sum of four terms $\mathcal{L}_2$ to $\mathcal{L}_5$. Defining with $X = -g_{\mu\nu}\phi^{,\mu}\phi^{,\nu}/2$ the canonical kinetic term, the four terms are specified by two non-canonical kinetic functions $K(\phi, X)$ and $G_3(\phi, X)$ and by two coupling functions $G_{4,5}(\phi, X)$, all of them in principle arbitrary:

$$S = \int d^4x\sqrt{-g}\sum_{i=2}^{5}\mathcal{L}_i + S_m \tag{5.108}$$

[a] Text adapted from Amendola et al. *Universe*, vol. 6, p. 20

where S_m is the action for matter fields – dark matter, baryons and radiation – and

$$\begin{aligned}\mathcal{L}_2 =& K(\phi, X)\,,\\ \mathcal{L}_3 =& -G_3(\phi, X)\Box\phi\,,\\ \mathcal{L}_4 =& G_4(\phi, X)R + G_{4,X}\left[(\Box\phi)^2 - (\nabla_\mu\nabla_\nu\phi)^2\right]\,,\\ \mathcal{L}_5 =& G_5(\phi, X)G_{\mu\nu}\nabla^\mu\nabla^\nu\phi - \frac{G_{5,X}}{6}\left[(\Box\phi)^3 - 3\,(\Box\phi)\,(\nabla_\mu\nabla_\nu\phi)^2 + 2\,(\nabla_\mu\nabla_\nu\phi)^3\right].\end{aligned} \tag{5.109}$$

(here $f_{,X} \equiv \partial f/\partial X$) Note that G_3 and G_5 must have an X dependence, otherwise they are total derivatives and could be rewritten – after integration by parts – as K and G_4 respectively. As usual, each term in the HL has dimension $mass^4$. Often one chooses the scalar field to have dimensions of mass, but this is not necessary. As already mentioned, the HL is the most general Lagrangian for a single scalar which gives second-order equations of motion for both the scalar and the metric on an arbitrary background. This is a necessary, but not sufficient, condition for the absence of instabilities. The terms $\mathcal{L}_4, \mathcal{L}_5$ couple the field ϕ to the Ricci scalar R and the Einstein tensor $G_{\mu\nu} = R_{\mu\nu} - Rg_{\mu\nu}/2$. As a consequence, $G_{4,5}$ are the gravity-modifying coupling function.

Let us now briefly discuss some useful limits of the HL.

- If $G_4 = 1/2$ and $G_5 = 0$ (it is actually sufficient $G_5 = const$) the HL reduces to ordinary gravity with a scalar field having a non-canonical kinetic sector given by $\mathcal{L}_2, \mathcal{L}_3$. The canonical form is obtained for $K = X - V(\phi)$ and $G_3 = 0$ ($G_3 = const$ is sufficient). ΛCDM is recovered for $K = -2\Lambda$.
- The "minimal" form of modified gravity within the HL is provided by $G_4 = G_4(\phi)$ and $G_5 = const$: this is then equivalent to a Brans-Dicke scalar-tensor model, again with a non-canonical kinetic sector.
- The original Brans-Dicke model is recovered assuming a kinetic sector, $K = (\omega_{BD}/\phi)X, G_3 = 0$, and $G_4(\phi) = \phi/2$.
- If, moreover, the kinetic sector vanishes, $K_{,X} = G_3 = 0$, then we reduce ourselves to a $f(R)$ model [388], whose Lagrangian is $\mathcal{L}_R = (R + f(R))/2$. In fact, this model is equivalent to a scalar-tensor theory with $G_4(\phi) = e^{2\phi/\sqrt{6}}/2$ and a potential $K(\phi) = -(Rf_{,R} - f)/2$ where $\phi = \sqrt{6}/2\log(1 + f_{,R})$. This relation should then be inverted to get $R = R(\phi)$ and used to replace R with ϕ in $K(\phi)$.
- If one sets $G_i(\phi, X) = G_i(X)$ then the Lagrangian is invariant under the shift $\phi \to \phi + c$ with $c = const$. This shift-symmetric version of the HL is connected to the Covariant Galileon when the functional dependence of the G_i is fixed and is able to produce the accelerated expansion without a potential that makes the field slow roll.

In general, the equations of motion for the scalar will couple it to the matter energy density. The full set of equations of motion has been studied in several papers, for instance in [418]. Any modification of the HL, or addition of terms (except the so-called Beyond Horndeski terms), based on the same scalar field, will introduce higher order equations of motion and associated instabilities, as a consequence of the Ostrogradsky theorem [419]. Of course one can in principle add several scalar fields, but on grounds of simplicity this is rather unnatural. Notice that we do not demand that the ϕ drives the present-day accelerated expansion. It could be, after all, that the modification of gravity and the accelerated expansion are independent phenomena. It would be very interesting, though, to explain the latter in terms of the former.

An interesting property of the HL is that, differently from the simple Brans-Dicke case, the velocity c_T of propagation of tensor modes (i.e. gravitational waves) is different from c. In fact, one has

$$c_T = 1 + \frac{2X\left(2G_{4X} - 2G_{5\phi} - \left(\ddot{\phi} - \dot{\phi}H\right)G_{5X}\right)}{2\left(G_4 - 2XG_{4X} + XG_{5\phi} - \dot{\phi}HXG_{5X}\right)} \tag{5.110}$$

The recent discovery of gravitational waves with optical counterparts has shown that $c_T = 1$ to very high precision. So $c_T = 1$ only if $G_4 = G_5 = 0$ or for particular (and very fine-tuned) combinations of G_4, G_5 (or, trivially, when $X = 0$, so for a static field). Of course, one can also devise models such that $c_T = 1$ today but not in the past.

5.7 Screening mechanisms

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 [319, 320]. 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. This effect is called *screening*.

An effective coupling $Q_{\rm eff}$ 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 [319, 320]. 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 $|Q_{\rm eff}|$ 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 details.

Let's start with the equations of the coupled dark energy

$$\nabla_\mu T^\mu_{\nu(\phi)} = -Q\,T_M \nabla_\nu \phi\,, \qquad \nabla_\mu T^\mu_{\nu(m)} = +Q\,T_m \nabla_\nu \phi\,, \tag{5.111}$$

where Q is taken constant. We have seen that the matter equation is solved by (see 4.22) $\rho_m = \rho e^{Q(\phi-\phi_0)}$, where $\rho=\rho_{m0}a^{-3}$. From now on we rescale the field do that $\phi_0=0$. The $\nu=0$ field equation is then

$$\Box\phi - V_{,\phi} = Q\rho e^{Q\phi}\,, \tag{5.112}$$

Eq. (5.112) can be written as

$$\Box\phi = V_{,\phi} + Q\rho e^{Q\phi}\,. \tag{5.113}$$

In a spherically symmetric spacetime of the Minkowski background (i.e. weak gravity background) this reduces to

$$\frac{{\rm d}^2\phi}{{\rm d}r^2} + \frac{2}{r}\frac{{\rm d}\phi}{{\rm d}r} = \frac{{\rm d}V_{\rm eff}(\phi)}{{\rm d}\phi}\,, \tag{5.114}$$

where r is the distance from the center of symmetry and the effective potential $V_{\rm eff}$ is defined by

$$V_{\rm eff}(\phi) \equiv V(\phi) + \rho e^{Q\phi}\,. \tag{5.115}$$

In a strong gravity background, Eq. (5.114) 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 assume now 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 $V_{\rm eff}$ has minima at field values ϕ_A and ϕ_B given by

$$V_{,\phi}(\phi_A) + Q\rho_A e^{Q\phi_A} = 0\,, \tag{5.116}$$

$$V_{,\phi}(\phi_B) + Q\rho_B e^{Q\phi_B} = 0\,. \tag{5.117}$$

The former corresponds to the region with a high density (interior of the body) with a large mass squared $m_A^2 \equiv \frac{{\rm d}^2V_{\rm eff}}{{\rm d}\phi^2}(\phi_A)$, whereas the latter to the lower density region (exterior of the body) with a small mass squared $m_B^2 \equiv \frac{{\rm d}^2V_{\rm eff}}{{\rm d}\phi^2}(\phi_B)$. See Fig. 5.3 for illustration.

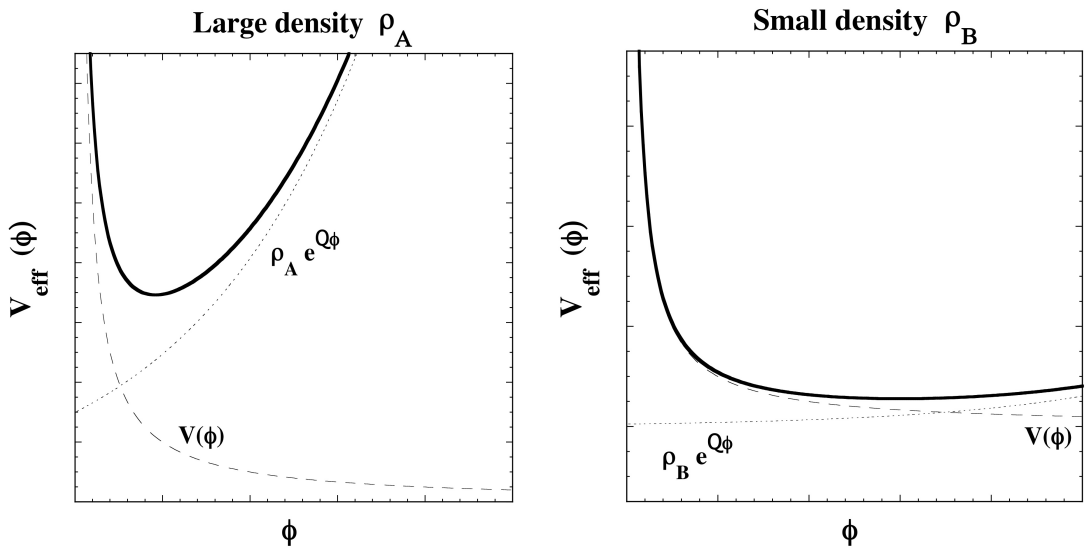

Figure 5.3: The effective potential $V_{\rm eff}$ of a chameleon field (solid curve) for the case $V_{,\phi} < 0$ and $Q > 0$. The effective potential is the sum of the field potential $V(\phi)$ (dashed curve) and the coupling term $\rho e^{Q\phi}$ (dotted curve). The left and right panels correspond to large and small densities, respectively. The field tends to be more massive around the minimum of the effective potential for larger density ρ.

Equation (5.114) shows that we need to consider the potential $(-V_{\rm eff})$ in order to find the "dynamics" of the field with respect to r. This means that the effective potential $(-V_{\rm eff})$ has a maximum at $\phi = \phi_A$, see Fig. 5.4. We impose the following boundary conditions:

$$\frac{{\rm d}\phi}{{\rm d}r}(r=0) = 0\,, \quad \phi(r \to \infty) = \phi_B\,. \tag{5.118}$$

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 $0 < r < r_1$. We say that the body has a thin-shell if r_1 is close to the radius r_c of the body. The value of r_1 will be determined later on. We split now the problem into three regions: region I for $0 < r < r_1$, region II for $r_1 < r < r_c$, and region III for $r > r_c$. We also assume that $Q\phi \ll 1$ because of constraints from local gravity.

In the region $0 < r < r_1$, the r.h.s. of Eq. (5.114) can be approximated as ${\rm d}V_{\rm eff}/{\rm d}\phi \simeq m_A^2(\phi - \phi_A)$ around $\phi = \phi_A$. Therefore Eq. (5.114) can be written as

$$\frac{{\rm d}^2\phi}{{\rm d}r^2} + \frac{2}{r}\frac{{\rm d}\phi}{{\rm d}r} = \frac{(r^2\phi')'}{r^2} = m_A^2(\phi - \phi_A)\,, \tag{5.119}$$

(prime is derivative wrt r). One can easily see by integrating the last equality (or by substitution) that the equation is solved by $\phi \sim e^{\pm m_A r}/r$ and also by $\phi = \phi_A = const.$ Hence we can write $\phi(r) = \phi_A + Ae^{-m_A r}/r + Be^{m_A r}/r$, where A and B are integration constants. To avoid the divergence of ϕ at $r = 0$ one needs $B = -A$, so the solution reduces to

$$\phi(r) = \phi_A + \frac{A(e^{-m_A r} - e^{m_A r})}{r} \qquad (0 < r < r_1)\,, \tag{5.120}$$

which satisfies the boundary condition $\frac{{\rm d}\phi}{{\rm d}r}(r=0) = 0$.

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 ${\rm d}V_{\rm eff}/{\rm d}\phi \simeq Q\rho_A$ in Eq. (5.114). We then find the following solution

$$\phi(r) = \frac{1}{6}Q\rho_A r^2 - \frac{C}{r} + D \qquad (r_1 < r < r_c), \tag{5.121}$$

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. (5.114) dominates over the r.h.s.. Then the solution outside the body satisfying the boundary condition $\phi(r \to \infty) = \phi_B$ is given by

$$\phi(r) = \phi_B + \frac{E}{r} \qquad (r > r_c). \tag{5.122}$$

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 (5.120), (5.121), and (5.122) at $r = r_1$ and $r = r_c$. We then obtain the following field profile [321]

$$\begin{aligned}
\phi(r) &= \phi_A - \frac{1}{m_A(e^{-m_A r_1} + e^{m_A r_1})}\left[\phi_B - \phi_A + \frac{1}{2}Q\rho_A(r_1^2 - r_c^2)\right]\frac{e^{-m_A r} - e^{m_A r}}{r}\,, \qquad (5.123)\\
&\quad (0 < r < r_1),\\
\phi(r) &= \phi_B + \frac{1}{6}Q\rho_A(r^2 - 3r_c^2) + \frac{Q\rho_A r_1^3}{3r}\\
&\quad - \left[1 + \frac{e^{-m_A r_1} - e^{m_A r_1}}{m_A r_1(e^{-m_A r_1} + e^{m_A r_1})}\right]\left[\phi_B - \phi_A + \frac{1}{2}Q\rho_A(r_1^2 - r_c^2)\right]\frac{r_1}{r}\,, \qquad (5.124)\\
&\quad (r_1 < r < r_c),\\
\phi(r) &= \phi_B - \left[r_1(\phi_B - \phi_A) + \frac{1}{6}Q\rho_A r_c^3\left(2 + \frac{r_1}{r_c}\right)\left(1 - \frac{r_1}{r_c}\right)^2\right.\\
&\quad \left. + \frac{e^{-m_A r_1} - e^{m_A r_1}}{m_A(e^{-m_A r_1} + e^{m_A r_1})}\left\{\phi_B - \phi_A + \frac{1}{2}Q\rho_A(r_1^2 - r_c^2)\right\}\right]\frac{1}{r}\,, \qquad (5.125)\\
&\quad (r > r_c)\,.
\end{aligned}$$

In the original papers of the chameleon mechanism [319, 320], 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. [321] this was revisited to match three solutions (5.120), (5.121), and (5.122) at $r = r_1$ and $r = r_c$.

The radius r_1 is determined by the condition that $dV_{\rm eff}/d\phi \approx m_A^2\,[\phi(r_1) - \phi_A]$ in region I equals $dV_{\rm eff}/d\phi \approx Q\rho_A e^{Q\phi_A} \approx Q\rho_A$ in region II. The condition $m_A^2\,[\phi(r_1) - \phi_A] = Q\rho_A$ translates into

$$\phi_B - \phi_A + \frac{1}{2}Q\rho_A(r_1^2 - r_c^2) = \frac{6Q\Phi_c}{(m_A r_c)^2}\frac{m_A r_1(e^{m_A r_1} + e^{-m_A r_1})}{e^{m_A r_1} - e^{-m_A r_1}}\,, \tag{5.126}$$

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

$$\phi(r) = \phi_B - 2Q_{\rm eff}\frac{GM_c}{r}\,, \tag{5.127}$$

where

$$Q_{\rm eff} = Q\left[1 - \frac{r_1^3}{r_c^3} + 3\frac{r_1}{r_c}\frac{1}{(m_A r_c)^2}\left\{\frac{m_A r_1(e^{m_A r_1} + e^{-m_A r_1})}{e^{m_A r_1} - e^{-m_A r_1}} - 1\right\}\right]. \tag{5.128}$$

The fifth force on a test particle of unit mass and a coupling Q is given by $\boldsymbol{F}_\phi = -Q\nabla\phi$. Hence its amplitude in the region $r > r_c$ is

$$F_\phi = 2\,|QQ_{\rm eff}|\,\frac{GM_c}{r^2}\,. \tag{5.129}$$

As long as $|Q_{\rm eff}| \ll 1$ it is possible to make the fifth force suppressed relative to the gravitational force GM_c/r^2. From Eq. (5.128) 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.

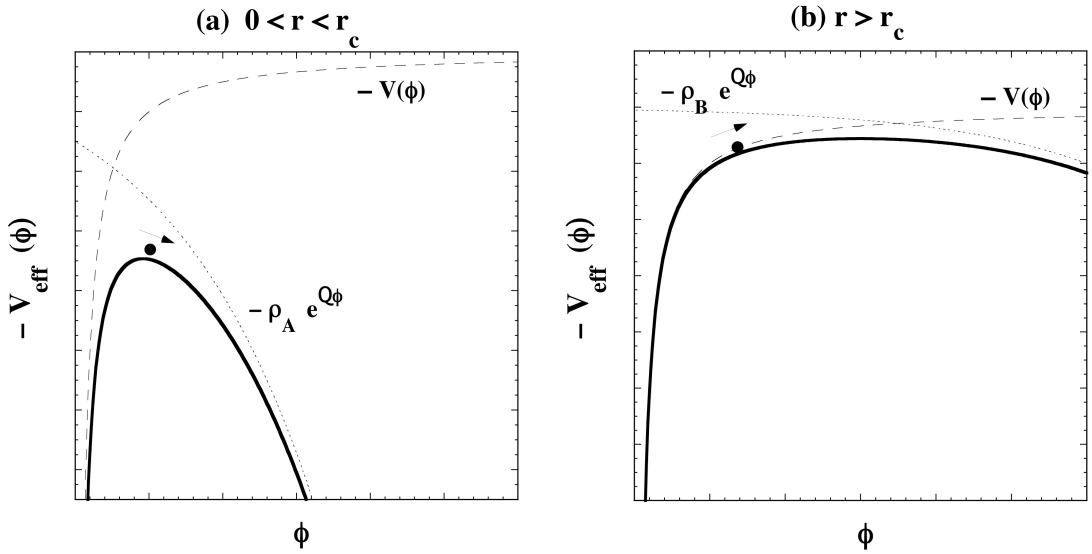

Figure 5.4: The inverted effective potential ($-V_{\rm eff}$) of a chameleon field inside (left) and outside (right) a spherically symmetric body with constant matter densities ρ_A and ρ_B, respectively. The black points represent the position of the field and show how the field "evolves" with the increase of r. When 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 ($r_1 < r < r_c$).

When the body has a thin-shell ($\Delta r_c \ll r_c$), one can expand Eq. (5.126) in terms of the small parameters $\Delta r_c/r_c$ and $1/(m_A r_c)$:

$$\epsilon_{\rm th} \simeq \frac{\Delta r_c}{r_c} + \frac{1}{m_A r_c}\,, \tag{5.130}$$

where $\epsilon_{\rm th}$ is the so called thin-shell parameter defined by

$$\epsilon_{\rm th} \equiv \frac{\phi_B - \phi_A}{6Q\Phi_c}\,. \tag{5.131}$$

As long as $m_A r_c \gg (\Delta r_c/r_c)^{-1}$, this recovers the relation $\epsilon_{\rm th} \simeq \Delta r_c/r_c$ [319, 320]. The effective coupling (5.128) is approximately given by

$$Q_{\rm eff} \simeq 3Q\epsilon_{\rm th}\,. \tag{5.132}$$

Under the condition $\epsilon_{\rm th} \ll 1$ one has $|Q_{\rm eff}| \ll |Q|$, which means that the smallness of the thin-shell parameter is crucially important for the compatibility with local gravity constraints. From Eqs. (5.123) and (5.124) 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 $r_1 \to 0$ in Eq. (5.128). One has $Q_{\rm eff} \simeq Q$ 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.

When the spherically symmetric body has a thin-shell, we can place experimental bounds on the thin-shell parameter $\epsilon_{\rm th}$ from the possible violation of the equivalence principle (EP). The tightest bound comes from the solar system tests of weak EP using the free-fall acceleration of Moon ($a_{\rm Moon}$) and Earth ($a_\oplus$) toward Sun [320]. The experimental bound on the difference of two accelerations is given by [323]

$$2\frac{|a_{\rm Moon} - a_\oplus|}{a_{\rm Moon} + a_\oplus} < 10^{-13}\,. \tag{5.133}$$

Applying this constraint to the thin-shell solution, we find that

$$\epsilon_{\mathrm{th},\oplus} < \frac{8.8 \times 10^{-7}}{|Q|} \,. \tag{5.134}$$

The constraint coming from the violation of strong EP provides a bound $\epsilon_{\mathrm{th},\oplus} \lesssim 10^{-4}$ [320], which is weaker than (5.134) for $|Q| = \mathcal{O}(1)$.

Chapter 6

Dark energy and linear cosmological perturbations[a]

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. For the basics of perturbation theory, see the *Lecture Notes on Cosmology.*

6.1 Perturbations for a general fluid

The standard linear perturbation equations for a single fluid and for two fluid cases that we have seen in the *Lecture Notes on Cosmology* can be generalized in several ways, such as considering more fluids, interaction terms, and various level 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 metric in $N = \ln a$:

$$\mathrm{d}s^2 = e^{2N}\left[-(1+2\Psi)\mathcal{H}^{-2}\mathrm{d}N^2 + (1+2\Phi)\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j\right] , \tag{6.1}$$

and work out from the beginning a single mode k, putting $\Phi(r,a) = \Phi(a)e^{i\boldsymbol{k}\cdot\boldsymbol{r}}$, $\Psi(r,a) = \Psi(a)e^{i\boldsymbol{k}\cdot\boldsymbol{r}}$ etc. This new metric forces a new definition of the first-order four-velocity:

$$u^\alpha \;=\; \frac{\mathrm{d}x^\alpha}{\mathrm{d}s} = \left\{\frac{\mathrm{d}N}{a(1+\Psi)\mathcal{H}^{-1}\mathrm{d}N}, \frac{\mathrm{d}x^i}{a\mathcal{H}^{-1}\mathrm{d}N}\right\} = \left\{\frac{\mathcal{H}}{a}(1-\Psi), \frac{v^i}{a}\right\} , \tag{6.2}$$

where of course $a = e^N$ and $a\mathrm{d}N = \mathrm{d}a$. It is convenient also to define a new velocity divergence:

$$\theta_{\mathrm{new}} = \frac{i\boldsymbol{k}\cdot\boldsymbol{v}}{\mathcal{H}} = \frac{\theta_{\mathrm{old}}}{\mathcal{H}} , \tag{6.3}$$

[a] Adapted from Amendola & Tsujikawa, *Dark Energy. Theory and Observations*, CUP 2010.

For simplicity we drop the subscript "new" but will remind the new definition when necessary.

All the equations can be converted into conformal-time equations by using the rules

$$\frac{\mathrm{d}}{\mathrm{d}N} = \frac{1}{\mathcal{H}}\frac{\mathrm{d}}{\mathrm{d}\eta}\,, \tag{6.4}$$

$$\frac{\mathrm{d}^2}{\mathrm{d}N^2} = \frac{1}{\mathcal{H}^2}\frac{\mathrm{d}^2}{\mathrm{d}\eta^2} - \frac{\mathrm{d}\mathcal{H}/\mathrm{d}\eta}{\mathcal{H}^2}\frac{\mathrm{d}}{\mathrm{d}\eta}\,, \tag{6.5}$$

and into ordinary time equations by the same rules and replacing $\eta \to t$ and $\mathcal{H} \to H$.

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

$$\delta T^0_0 = -\rho_t \delta_t \overset{=}{} - \sum_i \rho_i \delta_i \tag{6.6}$$

$$ik_j(\delta T^j_0) = -(1+w_{\rm eff})\rho_t\theta_t = -\sum_i (1+w_i)\rho_i\theta_i \tag{6.7}$$

$$\delta T^1_1 = \delta T^2_2 = \delta T^3_3 = c^2_{s,t}\rho_t\delta_t = \sum_i c^2_{s,i}\rho_i\delta_i\,, \tag{6.8}$$

where the total perturbation variables are given by

$$\delta_t = \sum_i \Omega_i\delta_i\,, \tag{6.9}$$

$$\theta_t = \sum_i \frac{1+w_i}{1+w_{\rm eff}}\Omega_i\theta_i\,, \tag{6.10}$$

together with the total equation of state and the sound speed

$$w_{\rm eff} = \frac{P_t}{\rho_t} = \sum_i \Omega_i w_i\,, \tag{6.11}$$

$$c^2_{s,t} = \frac{\sum_i c^2_{s,i}\Omega_i\delta_i}{\delta_t} = \frac{\sum_i c^2_{s,i}\Omega_i\delta_i}{\sum_i \Omega_i\delta_i}\,. \tag{6.12}$$

Recall that Ω_i is defined by $\Omega_i \equiv \rho_i/\rho_t$.

The total equation of state $w_{\rm eff}$ satisfies the following relation

$$\frac{\mathcal{H}'}{\mathcal{H}} = 1 + \frac{H'}{H} = -\frac{1}{2} - \frac{3}{2}w_{\rm eff}\,. \tag{6.13}$$

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

$$c^2_{s(a),t} = \frac{\dot{P}_i}{\dot{\rho}_i} = w_i - \frac{w_i'}{3(1+w_i)}\,. \tag{6.14}$$

Suppose now all components are barotropic, $c^2_{s,i} = \mathrm{d}P_i/\mathrm{d}\rho_i$. Under which condition is the total fluid barotropic? If we impose the adiabatic condition,

$$\frac{\delta\rho_i}{\rho_i'} = \frac{\delta\rho_j}{\rho_j'} \quad \to \quad \frac{\delta_i}{1+w_i} = \frac{\delta_j}{1+w_j}\,, \tag{6.15}$$

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. (6.12), we find

$$c^2_{s,t} = \frac{\sum_i c^2_{s,i}\Omega_i\delta_1(1+w_i)}{\sum_i \Omega_i\delta_1(1+w_i)} = \frac{\sum_i c^2_{s,i}\rho_i(1+w_i)}{\sum_i \rho_i(1+w_i)} = \frac{\sum_i(\mathrm{d}P_i/\mathrm{d}\rho_i)\dot{\rho}_i}{\sum_i \dot{\rho}_i} = \frac{\dot{P}_t}{\dot{\rho}_t} = w_{\rm eff} - \frac{w_{\rm eff}'}{3(1+w_{\rm eff})}\,. \tag{6.16}$$

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. 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 $P_t = P_t(\rho_t)$.

The perturbation equations for a generic perfect fluid with density contrast δ_X and velocity divergence θ_X are given by (see *Lecture Notes on Cosmology*)

$$\delta'_X = 3(w - c_s^2)\delta_X - (\theta_X + 3\Phi')(1+w)\,, \tag{6.17}$$

$$\theta'_X = \left(3w - 1 - \frac{w'}{1+w} - \frac{\mathcal{H}'}{\mathcal{H}}\right)\theta_X + \frac{c_s^2}{\hat{\lambda}^2(1+w)}\delta_X + \frac{\Psi}{\hat{\lambda}^2}\,. \tag{6.18}$$

Here we have introduced the quantity

$$\hat{\lambda} \equiv \mathcal{H}/k = aH/k\,, \tag{6.19}$$

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 $w \neq -1$. Perturbing Einstein's equations, we obtain the following equations

$$\Phi = 3\hat{\lambda}^2\left(\frac{1}{2}\delta_t + \Psi - \Phi'\right)\,, \tag{6.20}$$

$$\Phi' = \Psi - \frac{3}{2}\hat{\lambda}^2\theta_t(1 + w_{\rm eff})\,, \tag{6.21}$$

where we have used the background equation, $3\mathcal{H}^2 = 8\pi G a^2\rho_t$. For $w = -1$ 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 (6.17) and (6.18) can be also applied to the *total* component, replacing the subscript X for t and w, c_s for $w_{\rm eff}, c_{s,t}$. 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.

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

$$\Phi = \frac{3}{2}\hat{\lambda}^2[\delta_t + 3\hat{\lambda}^2\theta_t(1 + w_{\rm eff})]\,, \tag{6.22}$$

$$\Phi' = -\frac{3}{2}\hat{\lambda}^2[\delta_t + \theta_t(3\hat{\lambda}^2 + 1)(1 + w_{\rm eff})]\,. \tag{6.23}$$

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]

$$\Delta_t \equiv \delta_t + 3\hat{\lambda}^2\theta_t(1 + w_{\rm eff})\,, \tag{6.24}$$

in any gauge in which the $(0, i)$ elements 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 non-linearities, of redshift distortions). From Eq. (6.22) 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 of $\hat{\lambda} \ll 1$. When this limit is not respected (e.g., when discussing CMB, ISW, lensing), we stick with Φ and Ψ.

At small scales $\hat{\lambda} \ll 1$ and for vanishing σ, the equations reduce to

$$\delta_m'' + \frac{1}{2}(1 - 3w_{\rm eff})\delta_m' = \frac{\Phi}{\hat{\lambda}^2}\,, \tag{6.25}$$

$$\Phi = \frac{3}{2}\hat{\lambda}^2\delta_t = \frac{3}{2}\hat{\lambda}^2(\Omega_m\delta_m + \Omega_{\rm DE}\delta_{\rm DE})\,, \tag{6.26}$$

$$c_{s,t}^2 = c_{s,{\rm DE}}^2\left[1 - \frac{3\hat{\lambda}^2\Omega_m\delta_m}{2\Phi}\right]\,. \tag{6.27}$$

If dark energy does not cluster then we have $\delta_{\rm DE} = 0$, so that $\Phi = 3\hat{\lambda}^2\Omega_m\delta_m/2$ from Eq. (6.26). Equation (6.27) shows that in this case the total sound speed $c_{s,t}$ vanishes. From Eq. (6.25) it follows that

$$\delta_m'' + \frac{1}{2}(1 - 3\Omega_{\rm DE}w_{\rm DE})\delta_m' - \frac{3}{2}\Omega_m\delta_m = 0\,, \tag{6.28}$$

where we have used $w_{\rm eff} = \Omega_{\rm DE}w_{\rm DE}$. This equation is completely fixed by assigning a $w_{\rm DE}(a)$ given by the model and the present value $\Omega_m^{(0)} = 1 - \Omega_{\rm DE}^{(0)}$ from which $\Omega_m(a) = \Omega_m^{(0)}a^{-3}/[\Omega_m^{(0)}a^{-3} + (1 - \Omega_m^{(0)})a^{-3(1+\hat{w}_{\rm DE})}]$, with

$$\hat{w}_{\rm DE}(N) = \frac{1}{N}\int_0^N w_{\rm DE}(\tilde{N})\,{\rm d}\tilde{N}\,. \tag{6.29}$$

In some simple cases an analytical solution in terms of hypergeometric functions can be found [476]. It is often more useful however to work with an approximate solution. By using the growth rate parameter f defined as $f = \delta'/\delta$, Eq. (6.28) can be written in the form

$$f' + f^2 + \left[\frac{1}{2} - \frac{3}{2}w_{\rm DE}(1 - \Omega_m)\right]f = \frac{3}{2}\Omega_m\,. \tag{6.30}$$

By using the Friedmann equation $3H^2 = 8\pi G(\rho_m + \rho_{\rm DE})$ together with the continuity equation $\dot{\rho}_{\rm DE} + 3H(1 + w_{\rm DE})\rho_{\rm DE} = 0$, we obtain

$$\Omega_m' = 3w_{\rm DE}(1 - \Omega_m)\Omega_m\,. \tag{6.31}$$

Combining Eqs. (6.30) and (6.31), it follows that

$$3w_{\rm DE}\Omega_m(1 - \Omega_m)\frac{{\rm d}f}{{\rm d}\Omega_m} + \left[\frac{1}{2} - \frac{3}{2}w_{\rm DE}(1 - \Omega_m)\right]f + f^2 = \frac{3}{2}\Omega_m\,. \tag{6.32}$$

Substituting $f = \Omega_m^\gamma$ into Eq. (6.32), we find that [477]

$$3w_{\rm DE}\Omega_m(1 - \Omega_m)\,(\ln\Omega_m)\,\frac{{\rm d}\gamma}{{\rm d}\Omega_m} - 3w_{\rm DE}\left(\gamma - \frac{1}{2}\right)\Omega_m + \Omega_m^\gamma - \frac{3}{2}\Omega_m^{1-\gamma} + 3w_{\rm DE}\gamma - \frac{3}{2}w_{\rm DE} + \frac{1}{2} = 0\,. \tag{6.33}$$

If the variation of $w_{\rm DE}(z)$ is slow so that the condition $|{\rm d}w_{\rm DE}/{\rm d}\Omega_m| \ll 1/(1 - \Omega_m)$ is satisfied, we obtain the following estimate for γ [478]:

$$\gamma = \frac{3(1 - w_{\rm DE})}{5 - 6w_{\rm DE}} + \frac{3}{125}\frac{(1 - w_{\rm DE})(1 - 3w_{\rm DE}/2)}{(1 - 6w_{\rm DE}/5)^2(1 - 12w_{\rm DE}/5)}(1 - \Omega_m) + \mathcal{O}((1 - \Omega_m)^2)\,. \tag{6.34}$$

Note that the ΛCDM model corresponds to $\gamma \simeq 6/11 \simeq 0.545$.

Another fit for γ is provided by [479]

$$\gamma = 0.545 + 0.05[1 + w_{\rm DE}(z = 1)]\,. \tag{6.35}$$

In Fig. 6.1 we show the behavior of the perturbation growth and the comparison with the fit (6.35).

It is useful to remark that while γ does not depend strongly on $w_{\rm DE}$, the rate $f = \Omega_m^\gamma$ is significantly affected by $w_{\rm DE}$. At any given z, in fact, the dark energy component is more important for higher $w_{\rm DE}$ (assuming a

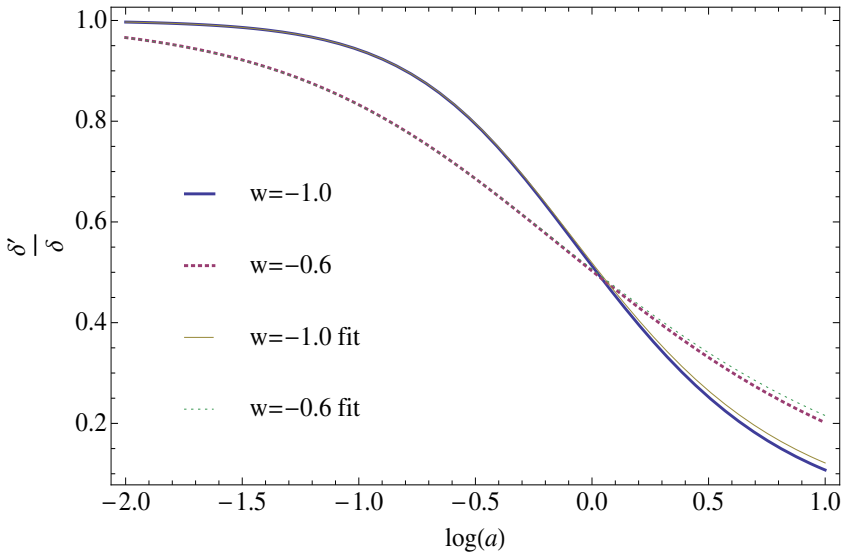

Figure 6.1: Growth rate $f = \delta_m'/\delta_m$ for $w_{\rm DE} = -1$ (full line) and $w_{\rm DE} = -0.6$ (dotted line) obtained by numerical integration compared to the approximation (6.35) (thin lines), fixing $\Omega_m^{(0)} = 0.3$.

constant $w_{\rm DE}$ for simplicity). Larger $\Omega_{\rm DE}$ means smaller Ω_m and hence $f(z)$ *decreases* with increasing $w_{\rm DE}$, 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 $w_{\rm DE}$. 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 $w_{\rm DE}$. 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 [480].

6.2 Perturbations of a scalar field

As we have seen, 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. [254, 482, 483, 484, 340].

Our dark energy model is therefore characterized by a general scalar-field potential $V(\phi)$ and general couplings $Q_i(\phi)$ to matter. The conservation equations with interacting terms for the field ϕ, cold dark matter (c), and baryons (b) are:

$$\nabla_\mu T^\mu_{(c)\nu} = Q_c(\phi) T_{(c)} \nabla_\nu \phi \,, \tag{6.36}$$

$$\nabla_\mu T^\mu_{(b)\nu} = Q_b(\phi) T_{(b)} \nabla_\nu \phi \,, \tag{6.37}$$

$$\nabla_\mu T^\mu_{(\phi)\nu} = -[Q_c(\phi) T_{(c)} + Q_b(\phi) T_{(b)}] \nabla_\nu \phi \,, \tag{6.38}$$

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 [319, 320] is at work. Here for generality we leave the couplings completely free.

As we have already seen in Sec. 4.1, the field equations in the flat FLRW background are given by

$$\ddot{\phi} + 3H\dot{\phi} + V_{,\phi} = -(Q_c\rho_c + Q_b\rho_b)\,, \tag{6.39}$$

$$\dot{\rho}_c + 3H\rho_c = Q_c\rho_c\dot{\phi}\,, \tag{6.40}$$

$$\dot{\rho}_b + 3H\rho_b = Q_b\rho_b\dot{\phi}\,, \tag{6.41}$$

$$\dot{\rho}_r + 4H\rho_r = 0\,, \tag{6.42}$$

$$3H^2 = \rho_\phi + \rho_c + \rho_b + \rho_r\,, \tag{6.43}$$

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 (6.1). In the following we neglect the contribution of radiation because it is unimportant when we discuss the perturbations after decoupling. If we define

$$u_\mu \equiv \frac{\phi_{,\mu}}{|g^{\alpha\beta}\Phi_{,\alpha}\phi_{,\beta}|^{1/2}}\,, \quad \rho_\phi \equiv -\frac{1}{2}g^{\alpha\beta}\phi_{,\alpha}\phi_{,\beta} + V\,, \quad P_\phi \equiv -\frac{1}{2}g^{\alpha\beta}\phi_{,\alpha}\phi_{,\beta} - V\,, \tag{6.44}$$

we realize that the energy-momentum tensor of the scalar field can be written as

$$T_{(\phi)\mu\nu} = (\rho_\phi + P_\phi)u_\mu u_\nu + g_{\mu\nu}P_\phi\,. \tag{6.45}$$

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 have H not $\mathcal{H}$)

$$\delta\rho_\phi = H^2(\phi'\varphi' - \phi'^2\Phi) + V_{,\phi}\varphi\,, \tag{6.46}$$

$$\delta P_\phi = H^2(\phi'\varphi' - \phi'^2\Phi) - V_{,\phi}\varphi\,, \tag{6.47}$$

$$\theta_\phi \equiv -\frac{ik_i(\delta T^i_{0(\phi)})}{(1+w_\phi)\rho_\phi} = \hat{\lambda}^{-2}\frac{\varphi}{\phi'}\,, \tag{6.48}$$

where $V_{,\phi} \equiv \mathrm{d}V/\mathrm{d}\phi$ and

$$\varphi \equiv \delta\phi\,, \tag{6.49}$$

denotes the field fluctuation. We also define the field density contrast

$$\delta_\phi \equiv \varphi/\phi\,. \tag{6.50}$$

The sound speed is therefore

$$c^2_{s,\phi} = \frac{\delta P_\phi}{\delta\rho_\phi} = \frac{H^2(\phi'\varphi' - \phi'^2\Phi) - V_{,\phi}\varphi}{H^2(\phi'\varphi' - \phi'^2\Phi) + V_{,\phi}\varphi}\,. \tag{6.51}$$

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 $\delta_\phi = 0$ 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

$$m^2_\phi \equiv \frac{\mathrm{d}^2V}{\mathrm{d}\phi^2}\,, \tag{6.52}$$

together with its dimensionless version

$$\hat{m}^2_\phi \equiv m^2_\phi/H^2\,. \tag{6.53}$$

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

$$\hat{V} \equiv V/H^2\,, \tag{6.54}$$

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

$$\delta_i' = 3(w_i - c_{s,i}^2)(1+3Q_i\phi')\delta_i - (\theta_i + 3\Phi')(1+w_i) + (1-3w_i)(Q_i\varphi' + Q_{i,\phi}\phi'\varphi)\,, \tag{6.55}$$

$$\theta_i' = -\frac{\theta_i}{2}\left[1 - 6w_i - 3w_{\rm eff} + \frac{2w_i'}{1+w_i} + 2Q_i(1-3w_i)\phi'\right] + \frac{1}{\hat{\lambda}^2}\left[\frac{c_{s,i}^2}{1+w_i}\delta_i + \Psi + \frac{Q_i(1-3w_i)}{1+w_i}\varphi\right], \tag{6.56}$$

where $Q_{i,\phi} \equiv {\rm d}Q_i/{\rm d}\phi$. Note that these equations reduce to Eqs. (6.17) and (6.18) in the limit $Q_i \to 0$. 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 $c_{s,i}^2 \to \delta P/\delta\rho = w - w'/[3(1+w)]$. 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

$$\varphi'' + \left(2 + \frac{\mathcal{H}'}{\mathcal{H}}\right)\varphi' + (\hat{\lambda}^{-2} + \hat{m}_\phi^2)\varphi - \phi'(3\Phi' - \Psi') + 2\hat{V}_{,\phi}\Psi$$
$$= -3\sum_i Q_i(1-3c_{s,i}^2)\Omega_i\delta_i - 6\sum_i Q_i(1-3w_i)\Omega_i\Psi - 3\sum_i(1-3w_i)Q_{i,\phi}\Omega_i\varphi\,. \tag{6.57}$$

This equation for $Q_i = 0$ can be obtained also directly from the continuity equation (6.55) by making use of the perturbation variables in Eqs. (6.46)-(6.48) and putting $\delta_i = \delta_\phi$.

Finally the equations for metric perturbations are

$$\Phi = \frac{\hat{\lambda}^2}{2}\left[2\hat{V}\Psi + \varphi\hat{V}_{,\phi} + \phi'(3\varphi + \varphi') + 3\sum\Omega_i\left\{\delta_i + 3(1+w_i)\hat{\lambda}^2\theta_i\right\}\right], \tag{6.58}$$

$$\Phi' = \frac{1}{2}\left[2\Psi - \varphi\phi' - 3\hat{\lambda}^2\sum(1+w_i)\theta_i\Omega_i\right], \tag{6.59}$$

plus the usual equivalence $\Phi = -\Psi$ in the absence of anisotropic stress.

Finally, in the uncoupled case with $\Phi = -\Psi$, the field perturbation equation (6.57) reduces to

$$\varphi'' + \left(2 + \frac{\mathcal{H}'}{\mathcal{H}}\right)\varphi' + (\hat{\lambda}^{-2} + \hat{m}_\phi^2)\varphi - 4\phi'\Phi' - 2\hat{V}_{,\phi}\Phi = 0\,. \tag{6.60}$$

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. (6.46), we obtain the following relation

$$\Omega_\phi\delta_\phi \simeq V_{,\phi}\varphi/(3H^2) \simeq -\phi'\varphi\,. \tag{6.61}$$

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_t 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 has to be analyzed separately.

6.3 From dark energy to dark force: the quasi-static limit

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

$$\Phi = \frac{\hat{\lambda}^2}{2}\left[3\sum\Omega_i\delta_i + \varphi\hat{V}_{,\phi} + \phi'(3\varphi + \varphi')\right], \tag{6.62}$$

$$\Phi' = -\frac{1}{2}\varphi\phi' - \Phi\,. \tag{6.63}$$

In the first equation we have used $\hat{\lambda}^2\hat{V} \lesssim \hat{\lambda}^2 \ll 1$. Inserting Eqs. (6.62) and (6.63) into Eq. (6.57) and taking the small-scale limit, we obtain

$$\varphi'' + \left(2 + \frac{\mathcal{H}'}{\mathcal{H}}\right)\varphi' + \left[\hat{\lambda}^{-2} + \hat{m}_\phi^2 + 2\phi'^2 + 3\sum(1-3w_i)\Omega_i Q_{i,\phi}\right]\varphi \simeq -3\sum Q_i(1-3c_{s,i}^2)\Omega_i\delta_i\,, \tag{6.64}$$

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. (6.64). Averaging over the rapid oscillations of φ, we obtain

$$\langle\varphi\rangle \simeq -3\hat{\lambda}^2\sum Q_i(1-3c_{s,i}^2)\Omega_i\delta_i\,. \tag{6.65}$$

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. (6.62) reduces to the usual Poisson equation

$$\Psi = -\frac{3}{2}\hat{\lambda}^2\sum_i \Omega_i\delta_i\,. \tag{6.66}$$

Now, if we substitute $\langle\varphi\rangle$ into Eq. (6.56), we can define a new potential acting on the j-th component (which includes the effect of the coupling Q_j)

$$\begin{aligned}\hat{\Psi}_j &\equiv \Psi + \frac{Q_j(1-3w_j)}{1+w_j}\langle\varphi\rangle \\ &= -\frac{3}{2}\hat{\lambda}^2\sum_i\Omega_i\delta_i\left[1 + 2Q_iQ_j\frac{(1-3c_{s,i}^2)(1-3w_j)}{1+w_j}\right].\end{aligned} \tag{6.67}$$

Assuming for instance two matter components, CDM and baryons (subscripts c, b), we have a new potential on CDM:

$$\hat{\Psi}_c = -\frac{3}{2}\hat{\lambda}^2\left[\Omega_b\delta_b\,(1+2Q_bQ_c) + \Omega_c\delta_c\,(1+2Q_c^2)\right]. \tag{6.68}$$

In real space, this equation becomes

$$\nabla^2\hat{\Psi}_c = 4\pi G_{bc}\rho_b\delta_b + 4\pi G_{cc}\rho_c\delta_c\,, \tag{6.69}$$

where we have defined

$$G_{ij} = G\gamma_{ij}\,, \quad \gamma_{ij} \equiv 1 + 2Q_iQ_j\,. \tag{6.70}$$

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. (6.55) and (6.56), we have

$$\delta_c' = -\theta_c\,, \tag{6.71}$$

$$\theta_c' = -\frac{1}{2}(1 - 3w_{\rm eff} + 2Q_c\phi')\theta_c + \hat{\lambda}^{-2}\hat{\Psi}_c\,, \tag{6.72}$$

$$\delta_b' = -\theta_b\,, \tag{6.73}$$

$$\theta_b' = -\frac{1}{2}(1 - 3w_{\rm eff} + 2Q_b\phi')\theta_b + \hat{\lambda}^{-2}\hat{\Psi}_b\,, \tag{6.74}$$

$$\hat{\Psi}_c = -\frac{3}{2}\hat{\lambda}^2(\gamma_{bc}\Omega_b\delta_b + \gamma_{cc}\Omega_c\delta_c)\,, \tag{6.75}$$

$$\hat{\Psi}_b = -\frac{3}{2}\hat{\lambda}^2(\gamma_{bb}\Omega_b\delta_b + \gamma_{bc}\Omega_c\delta_c)\,. \tag{6.76}$$

Differentiating Eq. (6.71) with respect to N and using Eq. (6.72), we obtain

$$\delta_c'' + \frac{1}{2}(1 - 3w_{\rm eff} + 2Q_c\phi')\delta_c' - \frac{3}{2}(\gamma_{cc}\delta_c\Omega_c + \gamma_{bc}\delta_b\Omega_b) = 0\,. \tag{6.77}$$

Similarly the equation for δ_b is given by

$$\delta_b'' + \frac{1}{2}(1 - 3w_{\rm eff} + 2Q_b\phi')\delta_b' - \frac{3}{2}(\gamma_{bb}\delta_b\Omega_b + \gamma_{bc}\delta_c\Omega_c) = 0\,. \tag{6.78}$$

These equations generalize the previous uncoupled equations.

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 $b =$ constant, we obtain the coupled equations

$$f' + f^2 + \frac{1}{2}(1 - 3w_{\rm eff} + 2Q_c\phi')f - \frac{3}{2}\gamma_{cc}\Omega_c = 0\,, \tag{6.79}$$

$$f' + f^2 + \frac{1}{2}(1 - 3w_{\rm eff} + 2Q_b\phi')f - \frac{3}{2b}\gamma_{bc}\Omega_c = 0\,, \tag{6.80}$$

from which by subtraction

$$b = \frac{3\gamma_{bc}\Omega_c}{3\gamma_{cc}\Omega_c - 2(Q_c - Q_b)\phi' f}\,. \tag{6.81}$$

Notice that all terms on the r.h.s. are in general function 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 (6.79). A simple analytical solution exists if $w_{\rm eff}, Q_c, \Omega_c, \phi'$ are constants (we are neglecting the baryons here):

$$f = -\frac{1}{4}(1 - 3w_{\rm eff} + 2Q_c\phi') \pm \frac{1}{4}\sqrt{(1 - 3w_{\rm eff} + 2Q_c\phi')^2 + 24\gamma_{cc}\Omega_c}\,. \tag{6.82}$$

This particular case occurs indeed on stationary solutions, e.g. the solution (d) in Section 4.1. In a pure matter cosmology we recover the standard solution $f = 1, -3/2$ for $w_{\rm eff} = Q_c = 0$ 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

$$\phi' = \sqrt{6}x_1 = 3/(Q_c + \lambda)\,. \tag{6.83}$$

Inserting the values of $\phi', w_{\rm eff}, \Omega_c$ for the point (d) into Eq. (6.82), we find that the growing mode solution corresponds to

$$f \simeq \sqrt{3Q_c\lambda}\,, \quad \text{for} \quad Q_c \gg \lambda. \tag{6.84}$$

This diverges for $Q_c \to \infty$, 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 $w_{\rm eff} \to -1$. 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 [485].

Another simple case is the ϕMDE scaling, i.e. the solution (a) in Sec. 4.1. Applying Eq. (6.82) 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, it is also possible to find approximate solutions in the traditional form $f = \Omega_m^\gamma$, where [486]

$$\gamma \approx 0.55(1 - 2.6Q_c^2)\,. \tag{6.85}$$

6.4 The Yukawa potential

Let us now consider the mass of the dark energy field, previously neglected. If $\hat{\lambda}^{-2}$ is not much larger than $\hat{m}^2 = \hat{m}_\phi^2$, Eq. (6.65) 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):

$$\varphi = -3Y(k)\hat{\lambda}^2 Q\Omega_m\delta_m\,, \tag{6.86}$$

where

$$Y(k) \equiv \frac{k^2}{k^2 + a^2m^2}\,, \tag{6.87}$$

and $m = \hat{m}H$. If we substitute Eq. (6.86) into Eq. (6.56), we find that the effective potential is given by (neglecting the baryons)

$$\hat{\Psi} = -\frac{3}{2}\hat{\lambda}^2\Omega_m\delta_m[1 + 2Q^2Y(k)]. \tag{6.88}$$

As we have seen in (4.22), in coupled DE the mass varies with the field ϕ as

$$M(\phi) = \rho_m a^3 = M_0 e^{\int Q\,\mathrm{d}\phi}\,. \tag{6.89}$$

Now, let us write down the present density in real space for a particle of mass M located at the origin as $\rho_m^{(0)} = M\delta_D(0)$. The density contrast in Fourier space is

$$\rho_m = M\int d^3x e^{-i\boldsymbol{k}\cdot\boldsymbol{x}}\delta_D(0) = M \tag{6.90}$$

(one could consider a factor of V^{-1} to make the density dimensionally correct, but the factor of V is irrelevant and would cancel later on, so one can consider a unitary volume). So in "empty" space, i.e. for $\rho_m \gg \rho_t$, we have:

$$\Omega_m\delta_m = \frac{\rho_m}{\rho_t} = \frac{\kappa^2 a^3\rho_m}{3\mathcal{H}^2 a} = \frac{\kappa^2 M(\phi)}{3\mathcal{H}^2 a}\,, \tag{6.91}$$

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)

$$\hat{\Psi}_k = -\frac{3}{2}\Omega_m\delta_m\hat{\lambda}^2[1 + 2Q^2Y(k)] = -4\pi GM(\phi)\left(\frac{1}{k^2} + \frac{2Q^2}{k^2 + a^2m^2}\right)\frac{1}{a}\,, \tag{6.92}$$

from which we can define an effective $G_{\rm eff}$ in Fourier space:

$$G_{\rm eff} = G\left(1 + \frac{2Q^2k^2}{k^2 + a^2m^2}\right)\,. \tag{6.93}$$

Under the Fourier transformation

$$\hat{\Psi}(x) = \frac{1}{(2\pi)^3}\int e^{i\boldsymbol{k}\cdot\boldsymbol{x}}\,\hat{\Psi}_k\,\mathrm{d}^3k\,, \tag{6.94}$$

we have, using the angular integral

$$\int e^{ikx\cos\theta}\,\sin\theta\mathrm{d}\theta\mathrm{d}\phi = 4\pi\frac{\sin kx}{kx}\,, \tag{6.95}$$

the result

$$\begin{aligned}\hat{\Psi}(x) &= -\frac{2GM(\phi)}{\pi a}\int_0^\infty \frac{\sin kx}{kx}\left(1 + \frac{2Q^2k^2}{k^2 + a^2m^2}\right)\mathrm{d}k \\ &= -\frac{2GM(\phi)}{\pi r}\int_0^\infty \frac{\sin y}{y}\left(1 + \frac{2Q^2y^2}{y^2 + m^2r^2}\right)\mathrm{d}y\,,\end{aligned} \tag{6.96}$$

where $\boldsymbol{r} = a\boldsymbol{x}$ is the physical coordinate. The last integral gives finally the Yukawa potential

$$\hat{\Psi}(r) = -\frac{GM(\phi)}{r}\left(1 + 2Q^2 e^{-mr}\right) . \tag{6.97}$$

that replaces Newton's potential in presence of a massive coupled scalar field. Notice that in general both Q and m can be function of ϕ and therefore of space and time. The same Yukawa correction applies to $f(R)$ gravity ($Q = -1/\sqrt{6}$). One sees therefore that the interaction range is given by m^{-1}.

When inserted in the perturbation equations, the Yukawa potential (6.92) will induce a k-dependence in the growth factor. This could be a way to test for modified gravity.

Chapter 7

Non-linear perturbations: simplified approaches

Quick summary

- Strongly non-linear fluctuations are difficult to handle and normally one has to employ powerful numerical simulations
- We first introduce the Zel'dovich approximation, which allows to follow in an almost analytical way the initial stages of structure formation beyond linearity
- Some more analytical results can be obtained assuming a spherical collapse. On scales of galaxies and clusters, Newtonian physics is sufficient
- Spherical collapse gives a simple but surprisingly accurate expression for the density of collapse and of virialization
- Using the so-called Press-Schechter formalism, one can approximately predict the number density of collapsed object as a function of their mass, to be compared to real data or simulations
- In this entire chapter we can safey use Newtonian gravity since we deal with scales well smaller than the horizon.

7.1 The Zel'dovich approximation

So far we have only investigated linear perturbations. Stars, galaxies and clusters, however, are certainly not linear objects. For instance, the density contrast of a typical cluster of galaxies can be $\delta > 200$. Going from the linear treatment to the non-linear one is however generally very difficult. Even if some important step forward can be achieved by going to higher order in perturbation theory, ultimately one needs large N-body simulations.

A popular way to make progress in non-linear evolution before ressorting to numerical methods is to adopt the Zel'dovich approximation. The idea is to follow the movement of particles under the action of gravity until they hit each other and create a (fomally) infinite density. This should approximate the behavior of particles in a N-body simulation at least at some early time. Consider two sets of *comoving* coordinates. One, x_0, represents the coordinates of particles in an unperturbed Universe. Since they are comoving, they do not depend on time. The other, $\mathbf{x}(t)$, in a perturbed one. Initially, we perturb the position of each particle by a vector field $\mathbf{s}$, called displacement. Then we assume that at some later time t the position of the particles is given by

$$\mathbf{x}(t) = \mathbf{x}_0 + g(t)\mathbf{s}(\mathbf{x}_0) \tag{7.1}$$

This means we assume the position at time t only depends on the initial displacement through a time (and not space) dependent function, still to be defined. The density of the particle at any given time is $\rho(x, t)$ in the

perturbed Universe and $\rho_0(t)$ in the unperturbed one (which just follow the cosmic expansion, $\rho_0 \sim a^{-3}$). Since the particle number density $dn = \rho dV$ must be conserved, we have

$$\rho(x,t)d^3x = \rho_0(t)d^3x_0 \tag{7.2}$$

which implies

$$\rho(x,t) = \rho_0(t)\left|\frac{\partial \mathbf{x}}{\partial \mathbf{x}_0}\right|^{-1} \tag{7.3}$$

Let us assume now, without loss of generality, that the coordinates have been chosen along the direction of the eigenvectors of the deformation tensor

$$d_{ij} \equiv -\frac{\partial s_i}{\partial x_{0,j}} \tag{7.4}$$

In this case d_{ij} is diagonalized and therefore

$$\left|\frac{\partial \mathbf{x}}{\partial \mathbf{x}_0}\right| = \left|I + g(t)\frac{\partial \mathbf{s}(\mathbf{x}_0)}{\partial \mathbf{x}_0}\right| = |\delta_{ij} - g(t)d_{ij}| = (1-g\lambda_1)(1-g\lambda_2)(1-g\lambda_3) \tag{7.5}$$

where I and δ_{ij} represent the identity matrix and λ_i are the three eigenvalues of d_{ij} (we show below that the eigenvalues are real). This means

$$\rho(x,t) = \frac{\rho_0(t)}{(1-g\lambda_1)(1-g\lambda_2)(1-g\lambda_3)} \tag{7.6}$$

Before we comment on this important expression, let us understand the meaning of g and s. Expanding (7.6) for small $g\lambda_i$, we find

$$\rho(x,t) \approx \rho_0(t)(1 + g(t)(\lambda_1+\lambda_2+\lambda_3) = \rho_0(t)(1 + g(t)\mathrm{Tr}(d_{ij})) \tag{7.7}$$

and therefore

$$\delta(t) \equiv \frac{\rho(x,t)-\rho_0(t)}{\rho_0(t)} = -g(t)\frac{\partial s_i}{\partial x_{0,i}} = -g(t)\nabla_{x_0}\mathbf{s}(x_0) \tag{7.8}$$

This expression, being a linearized one, must coincide with the growth law, $\delta(t) = G(t)\delta_0$, where $G(t)$ is the growth function we have already evaluated for various cases. Then we see that we should identify $g(t)$ with $G(t)$ and

$$-\nabla_{x_0}\mathbf{s}(x_0) = \delta_0 \tag{7.9}$$

Now, from the Poisson equation and the Friedmann equation we have

$$\nabla^2\Psi = 4\pi\rho_m\delta = \frac{3}{2}a^2H^2\Omega_m G(t)\delta_0 \tag{7.10}$$

where the factor a^2 arises because we are adopting comoving, rather than physical, coordinates. Then we see that

$$\delta_0 = \frac{2}{3a^2H^2\Omega_m G}\nabla^2\Psi \tag{7.11}$$

and therefore

$$\mathbf{s}(x_0) = -\frac{2}{3a^2H^2\Omega_m G}\nabla\Psi \tag{7.12}$$

With this identification of $\mathbf{s}$, the deformation tensor d_{ij} is symmetric and therefore its three eigenvalues are real. Therefore, we have completely specified the prescription (7.1): $g(t)$ is the growth factor, and the initial

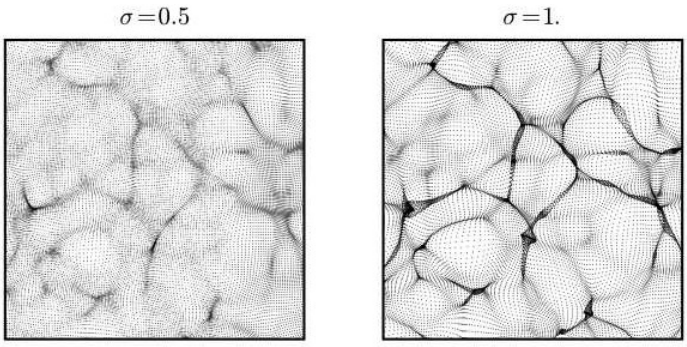

Figure 7.1: Formation of pancakes in a simulation based on the Zel'dovich approximation (from S. Shandarin, arXiv:0912.4520).

displacement field $\mathbf{s}$ is essentially the gradient of the gravitational potential, i.e. the force acting on the particles. In this way, one can run a very cheap N-body simulation: first, take the *linear* power spectrum at some early epoch for the model you want to simulate; second, convert the power spectrum for δ into a power spectrum for Ψ using Poisson equation in Fourier space; third, create a real space realization of this spectrum by overimposing sinusoidal oscillations with amplitude given by the spectrum and random phases; fourth, put particles on a regular grid; fifth, evaluate the displacement field by evaluating at every grid point (7.12); finally, move the particles out of their initial grid point by using (7.1).

To appreciate strenghts and limits of this technique, let us now come back to Eq. (7.6). Since $g(t)$ is a growing function (we discard the decaying mode, if any), $\rho(x,t)$ will develop a singularity as soon as one the largest λ_i is positive. This means that the particle will move primarily along the eigenvector associate to $\max\lambda_i$ and form regions of high density on the plane orthogonal to this direction: in other words, particle will tend to form planar structures, called pancakes (or *blinis* in the original Russian) by Zel'dovich, clearly visible in Fig. (7.1). After this singularity, the approximation will no longer be valid. In reality, is already quite surprising that the prescription (7.1) holds quite well beyond the linear regime!

Once the pancakes have been reached, one might assume that the particles "stick" onto, or oscillate around, the planar regions by friction or some hydrodynamic mechanism, and then continue flowing along the planes reaching the edges (called filaments) and finally slide along the filaments towards halos from which, in turn, galaxies and clusters will form. This is indeed qualitatively what is seen in full N-body simulations. Most of the current codes actually exploit the Zel'dovich approximation to speed up the calculations during the earliest stages at $z \gg 1$.

7.2 Spherical collapse[a]

After the formation of pancakes, the Zel'dovich approximation is no longer viable, although it can be extended through second-order schemes or *ad hoc* prescriptions. There is however a way to get, on a first approximation which however turns out to be surprisingly accurate, an estimate of an important observable, namely how many objects form (i.e., collapse into a virialized structure) for a given mass. This aproximation relies on sphericality and Gaussianity. The idea is first to find the value of the density contrast in the linear approximation at which a spherical perturbation collapse and virializes and then, find the fraction of the Gaussian distribution of perturbations that are above this δ collapse value. This fraction corresponds to the fraction of perturbations that form structures for a given mass.

For scales at which Newtonian thery applies, 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

$$\frac{\mathrm{d}^2R}{\mathrm{d}t^2} = -\frac{GM(R)}{R^2} = -\frac{4}{3}\pi G\rho R\,, \tag{7.13}$$

where $M(R) = 4\pi\rho R^3/3$ is the *constant* mass inside the shell. Since for pressureless matter the background

[a]Adapted from Amendola & Tsujikawa, *Dark Energy. Theory and Observations*, CUP 2010.

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

$$\delta = \left(\frac{a(t)R_0}{R}\right)^3 - 1\,, \tag{7.14}$$

inside the shell and $\delta = 0$ outside. The crucial assumption here is that δ is a step, or *top-hat*, function, which allows in fact to cancel all spatial derivatives. Replacing R with δ, the equation for δ in our time variable N is then:

$$\delta'' + \left(1 + \frac{\mathcal{H}'}{\mathcal{H}}\right)\delta' - \frac{3}{2}\Omega_m\delta = \frac{4}{3}\frac{\delta'^2}{1+\delta} + \frac{3}{2}\Omega_m\delta^2\,. \tag{7.15}$$

Multyiplying Eq. (7.13) on both sides by $2\mathrm{d}R/\mathrm{d}t$ the equation can be integrated once as

$$\left(\frac{\mathrm{d}R}{\mathrm{d}t}\right)^2 = \frac{2GM}{R} - C\,, \tag{7.16}$$

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:

$$\delta = \frac{9}{2}\frac{(\tau-\sin\tau)^2}{(1-\cos\tau)^3} - 1\,, \tag{7.17}$$

$$\delta_L = \frac{3}{5}\left[\frac{3}{4}(\tau-\sin\tau)\right]^{2/3}\,, \tag{7.18}$$

where δ_L (> 0) is the solution of the linearized equation, i.e. the left-hand-side of Eq. (7.15), normalized so that for $\tau \to 0$, $\delta_L \approx \delta$. Note that a_0, t_0 are chosen so that $\delta(\tau = 0) = 0$. 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 $\delta_{\rm coll}$ 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 [516, 517] 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. (7.18) 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

$$\delta_L = \delta_{\rm coll} = (3/5)(3\pi/2)^{2/3} \approx 1.686\,, \tag{7.19}$$

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, $\delta_{\rm coll}$ depends on time. An approximation for dark energy with constant $w_{\rm DE}$ in flat space is (Weinberg and Kamionkowski, MNRAS 341, 2003, 251)

$$\delta_{\rm coll}(z) = 1.686\,[1 + \alpha(w_{\rm DE})\log_{10}\Omega_m(z)]\,, \tag{7.20}$$

$$\alpha(w_{\rm DE}) = 0.353 w_{\rm DE}^4 + 1.044 w_{\rm DE}^3 + 1.128 w_{\rm DE}^2 + 0.555 w_{\rm DE} + 0.131\,. \tag{7.21}$$

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

$$K = \frac{R}{2}\frac{\partial U}{\partial R}\,. \tag{7.22}$$

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.

$$U_T = U_V + K_V = U_V/2\,. \tag{7.23}$$

Since for a uniform sphere $U = -3GM/5R$ (and remembering once again we are assuming $M =$ 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. A numerical fit for $w_{\rm DE} =$ constant models in flat space gives (Weinberg and Kamionkowski, MNRAS 341, 2003, 251)

$$\delta_V \approx 178[1 + b_1\theta^{b_2}(z)]\,, \tag{7.24}$$

$$\theta = \frac{1-\Omega_m(z)}{\Omega_m(z)}\,, \tag{7.25}$$

$$b_1 = 0.399 - 1.309(|w_{\rm DE}|^{0.426} - 1)\,, \tag{7.26}$$

$$b_2 = 0.941 - 0.205(|w_{\rm DE}|^{0.938} - 1)\,, \tag{7.27}$$

if z is the collapse redshift.

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.

7.3 The mass function of collapsed objects[b]

The main reason why it is worthwhile to discuss the abstract phenomenon as a "spherical collapse" is that the critical value $\delta_{\rm coll}$ and the virial radius R_V (or rather the mass contained within that radius) enter the Press-Schecther (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 $\delta_{\rm coll}$.

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 $\delta > \delta_{\rm coll}$ 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

$$p(M,z)|_{\delta>\delta_{\rm coll}} = \frac{1}{\sigma_M(z)\sqrt{2\pi}}\int_{\delta_{\rm coll}}^{\infty}\exp\left(-\frac{\delta_M^2}{2\sigma_M^2(z)}\right)\mathrm{d}\delta_M = \frac{1}{2}\,\mathrm{erfc}\left(\frac{\delta_{\rm coll}}{\sqrt{2}\sigma_M(z)}\right)\,, \tag{7.28}$$

where $\mathrm{erfc}(x)$ is the error function. The fraction containing objects of mass within the range $[M, M+\mathrm{d}M]$ is given by

$$\mathrm{d}p(M,z) = \left|\frac{\partial p(M,z)|_{\delta>\delta_{\rm coll}}}{\partial M}\right|\mathrm{d}M\,. \tag{7.29}$$

Remember that in general the threshold $\delta_{\rm coll}$ depends on z. Although the boxes with $\delta > \delta_{\rm coll}$ are certainly not in the linear regime, the idea is to use the linear regime to estimate the fraction of collapsed regions. We

[b] Adapted from Amendola & Tsujikawa, *Dark Energy. Theory and Observations*, CUP 2010.

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 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 $\mathrm{d}p$ of V, i.e.

$$NV_M = V\mathrm{d}p\,, \tag{7.30}$$

and therefore the number density $\mathrm{d}n$ of collapsed halos with mass in the $\mathrm{d}M$ range (the *mass function*) will be

$$\mathrm{d}n = \frac{N}{V} = \frac{\mathrm{d}p}{V_M} = \frac{\rho}{M}\left|\frac{\partial p(M,z)|_{\delta>\delta_{\rm coll}}}{\partial M}\right|\mathrm{d}M = \sqrt{\frac{2}{\pi}}\frac{\rho}{M^2}\frac{\delta_{\rm coll}}{\sigma_M}\left|\frac{\mathrm{d}\,\ln\sigma_M}{\mathrm{d}\,\ln M}\right| e^{-\delta^2_{\rm coll}/(2\sigma^2_M)}\mathrm{d}M\,. \tag{7.31}$$

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

$$V\int_0^\infty \left(\frac{\mathrm{d}n}{\mathrm{d}M}\right)\mathrm{d}M = 1\,. \tag{7.32}$$

This factor-of-2 adjustment can be justified with a random walk analysis of fluctuations. 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. (7.31) is sometimes written as

$$\frac{M}{\rho}\left|\frac{\mathrm{d}n}{\mathrm{d}\,\ln\sigma_M}\right| = f(\sigma_M, z)\,, \tag{7.33}$$

where all the cosmological information is contained in the function

$$f(\sigma_M, z) = \sqrt{\frac{2}{\pi}}\frac{\delta_{\rm coll}}{\sigma_M}e^{-\delta^2_{\rm coll}/(2\sigma^2_M)}\,. \tag{7.34}$$

The number density $\mathrm{d}n(M,z)$ can then be "directly" confronted with the observed densities of objects (clusters, galaxies, quasars) at any redshift. The mass M is often taken to be the virial mass of that class of objects. Because of the exponential dependence on $\delta_{\rm coll}/\sigma_M$, the PS formula is quite sensitive to the cosmological model (see Fig. 7.2).

The theoretical prediction depends then on the variance σ^2_M at any given epoch and mass and on the collapse threshold δ_c, which however is found to be only mildly dependent on cosmology. The variance σ^2_M can be written in linear growth approximation as $\sigma^2_M(z) = G^2(z)\sigma^2_M(0)$ where $G(z)$ is the growth function. It can be compared to observations once we have a reliable estimator of the cluster masses, for instance using the X-ray temperature scaling Eq. (10.7) or through the Sunyaev-Zel'dovich effect. This gives then a test of the power spectrum normalization σ^2_M. Since also Ω_m enters the growth function $G(z)$, the test constrains some combination of σ_8 and Ω_m.

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 absence of processes like merging, dissipation, shell crossing. Surprisingly, this shaky foundation did not prevent the PS formula to prove 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 or merging or by directly fitting to large N-body simulations. A remarkably successful fit is given by (Jenkins et al. MNRAS 321 (2001) 372)

$$f(\sigma_M, z) = 0.315\exp(-|0.61 - \ \ln\sigma_M(z)|^{3.8})\,. \tag{7.35}$$

This fit has been found to hold for a large range of masses, redshifts, and cosmological parameters, including dark energy with constant or varying $w_{\rm DE}$.

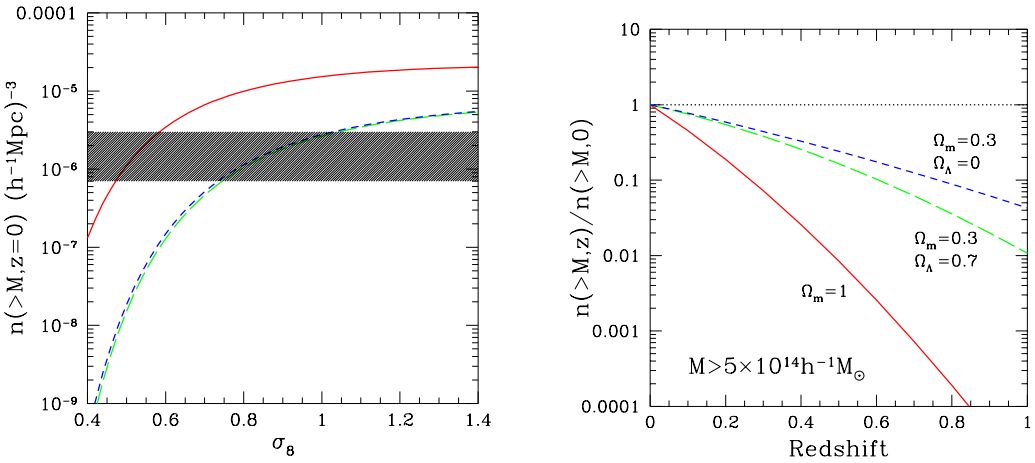

Figure 7.2: The sensitivity of the cluster mass function to cosmological models. Left panel: The cumulative mass function at $z = 0$ for $M > 5 \times 10^{14}\, h^{-1} M_\odot$ ($M_\odot$ is the solar mass) for three cosmologies, as a function of σ_8; solid line: $\Omega_m^{(0)} = 1$; short–dashed line: $\Omega_m^{(0)} = 0.3$, $\Omega_\Lambda^{(0)} = 0.7$; 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 Rosati, Borgani, Norman, ARAA 40 (2002) 539.

Chapter 8

Standard non-linear perturbation theory

In this chapter we introduce a systematic way to estimate the effect of non-linearities on the power spectrum. We work entirely in Newtonian gravity since we refer to scales much smaller than the horizon. A useful review for this section is Bernardeau et al. 2002. For the biased tracers, I followed Chudaykin et al (2020).

8.1 Second-order perturbations

The Newtonian fluidodynamical equations are (see *Lectures notes on Cosmology*)

$$\dot{\rho} + \mathbf{v} \cdot \nabla \rho = -\rho \nabla \cdot \mathbf{v} \qquad \text{conservation} \tag{8.1}$$

$$\rho(\dot{\mathbf{v}} + \mathbf{v} \cdot \nabla \mathbf{v}) = -\nabla p - \rho \nabla \Phi \qquad \text{Euler} \tag{8.2}$$

$$\nabla^2 \Phi = 4\pi\rho \qquad \text{Poisson} \tag{8.3}$$

where the dot is here a derivative wrt cosmic time, $\mathbf{v} = \mathbf{v}_p + H\mathbf{x}$ is the total velocity, including the Hubble expansion $\mathbf{v}_H = H\mathbf{x}$, and $\mathbf{v}_p$ the peculiar velocity. By introducing the density contrast $\delta = (\rho(x,t) - \rho_0(t))/\rho_0(t)$ (where ρ_0 is the background density that depends only on time) the first one can be written as ($\dot{\rho} = (\dot{\rho_0}\delta) + \dot{\rho}_0 = \dot{\rho_0}(1+\delta) + \rho_0\dot{\delta}$)

$$\dot{\rho_0}(1+\delta) + \rho_0\dot{\delta} + \rho_0(\mathbf{v_p} + H\mathbf{x}) \cdot \nabla\delta = -\rho_0(1+\delta)(\nabla \cdot \mathbf{v_p} + 3H) \tag{8.4}$$

Since at the background level $\dot{\rho}_0 + 3H\rho_0 = 0$, this becomes (we suppress the subscript p from now on)

$$\frac{d\delta}{dt} \equiv \dot{\delta} + H\mathbf{x} \cdot \nabla\delta = -\nabla \cdot (1+\delta)\mathbf{v} \tag{8.5}$$

On the rhs a nabla operator wrt physical coordinates $\mathbf{x} = a\mathbf{r}$, which becomes $a^{-1}\nabla_r$ converting to comoving coordinates. On the lhs, one can show that

$$\left(\frac{d\delta}{dt}\right)_x \equiv \dot{\delta} + \mathbf{v}_0 \cdot \nabla\delta = \left(\frac{\partial\delta}{\partial t}\right)_{\mathbf{r}} \tag{8.6}$$

(see proof in *Lectures notes on Cosmology).* Finally, therefore, the continuity equation is (dot is from now on the derivative wrt conformal time, so the a^{-1} factors cancel out)

$$\dot{\delta} = -\nabla(1+\delta)\mathbf{v} \tag{8.7}$$

where the nabla is now wrt comoving coordinates and we suppressed the subscript r.

Taking now the gradient of the Euler equation for a pressureless fluid, inserting the Poisson equation, we obtain

$$\nabla\dot{\mathbf{v}} + \nabla \cdot (\mathbf{v} \cdot \nabla\mathbf{v}) = -4\pi\rho \tag{8.8}$$

and steps analogous to above lead to

$$\dot{\theta} + H\theta - \frac{3}{2}a^2H^2\delta = -\nabla(\mathbf{v}\cdot\nabla\mathbf{v}) \tag{8.9}$$

where now $\mathbf{v}$ is the peculiar velocity, we moved to comoving coordinates, and we employed the Friedmann equation $3H^2 = 8\pi\rho$.

Eqs. (8.7) and (8.9) can be combined into a single non-linear second order equation if we assume that $\mathbf{v}$ is irrotational also at second order, i.e. $\mathbf{v} = \nabla w$, where w is a velocity potential and of course $\theta = \nabla^2 w$. Let us first rewrite the equations using $\log a$ as independent variable and redefining $\theta_{new} = \theta_{old}/\mathcal{H}$:

$$\frac{d\delta}{dN} \equiv \delta' + \mathbf{v}\nabla\delta = -(1+\delta)\theta \tag{8.10}$$

$$\theta' = -(1+\frac{\mathcal{H}'}{\mathcal{H}})\theta + \frac{3}{2}\delta - \nabla_i(v_j\nabla^j v^i) \tag{8.11}$$

The last term is

$$\nabla_i(v_j\nabla^j v^i) = \nabla_i\left[(\nabla_j w)\nabla^j\nabla^i w)\right] = (\nabla_i\nabla_j w)(\nabla^j\nabla^i w) + \left[(\nabla_j w)\nabla_i\nabla^j\nabla^i w\right] \tag{8.12}$$

$$= (\nabla_i\nabla_j w)(\nabla^j\nabla^i w) + \left[(\nabla_j w)\nabla^j\nabla^2 w)\right] = (\nabla_i\nabla_j w)^2 + v_i\nabla^i\theta \tag{8.13}$$

so

$$\frac{d\theta}{dN} \equiv \theta' + v_i\nabla^i\theta = -(1+\frac{\mathcal{H}'}{\mathcal{H}})\theta + \frac{3}{2}\delta - (\nabla_i\nabla_j w)^2 \tag{8.14}$$

Finally, in order to obtain a closed set of equations, we assume that the dominant component of the velocity is purely radial, i.e. the non-linear collapse is approximately spherical, at least initially. Then one can write $\mathbf{v} = v\{1,1,1\}/\sqrt{3}$ and

$$(\nabla_i\nabla_j w)^2 = (\nabla_i v_j)(\nabla^i v^j) = \frac{1}{3}\theta^2 \tag{8.15}$$

so finally

$$\frac{d\theta}{dN} \equiv \theta' + v_i\nabla^i\theta = -(1+\frac{\mathcal{H}'}{\mathcal{H}})\theta + \frac{3}{2}\delta - \frac{1}{3}\theta^2 \tag{8.16}$$

Now, differentiating (8.10) and inserting (8.16), and neglecting the gradient terms $\nabla\delta, \nabla\theta$ because near the center of a spherical perturbations they must indeed vanish, we obtain the same spherical collapse equation already encountered in Eq. (7.15):

$$\delta'' + \left(1 + \frac{\mathcal{H}'}{\mathcal{H}}\right)\delta' - \frac{3}{2}\Omega_m\delta = \frac{4}{3}\frac{\delta'^2}{1+\delta} + \frac{3}{2}\Omega_m\delta^2\,. \tag{8.17}$$

where now δ depends only on time. In a Einstein-de Sitter Universe in which $\Omega_m = 1$, one has $\frac{\mathcal{H}'}{\mathcal{H}} = -\frac{1}{2}$ and, as we know already, one finds at first order $G^{(1)} = G^{(1)\prime} = G^{(1)\prime\prime} = a$ and the growth function $f = 1$. At second order, we can neglect the denominator $1+\delta$ at rhs, and write

$$\frac{4}{3}\frac{\delta'^2}{1+\delta} + \frac{3}{2}\Omega_m\delta^2 \approx \left(\frac{4}{3}f^2 + \frac{3}{2}\Omega_m\right)\delta^2 = \frac{17}{6}\delta^2 \tag{8.18}$$

We now expand $\delta = G^{(1)}\delta^{(1)} + G^{(2)}(\delta^{(1)})^2 + ...$ Therefore at second order we get in EdS

$$G^{(2)\prime\prime} + \frac{1}{2}G^{(2)\prime} - \frac{3}{2}G^{(2)} = \frac{17}{6}G^{(1)2}\,. \tag{8.19}$$

Making the Ansatz $G^{(2)} = \alpha G^{(1)2}$ we find

$$G^{(2)} = \frac{17}{21}G^{(1)2}\,. \tag{8.20}$$

One sees therefore that the second-order perturbations grow as $G^{(1)2}$ in a EdS model. In the following we will assume therefore that the second order perturbations grow as $G^{(1)2}$.

8.2 Fourier space

We now consider in detail the *space* dependence of the second-order perturbations, in order to derive a correction to the power spectrum.

In the linear regime, Eq. (8.7) gives $\dot{\delta} = -\nabla \mathbf{v}$, which can be immediately Fourier-transformed as $\dot{\delta} = -i\mathbf{k}\mathbf{v} = -\theta$, so

$$\mathbf{v} = i\mathcal{H}\delta_k f \frac{\mathbf{k}}{k^2} \tag{8.21}$$

where δ_k is the total matter field. This can also be written as

$$\mathbf{v} = -i\theta \frac{\mathbf{k}}{k^2} \tag{8.22}$$

so we can convert $\mathbf{v}, \delta, \theta$ at linear level one into the other.

In Fourier space, we have $\delta = \int \delta_k e^{i\mathbf{k}\mathbf{x}} d^3k$, and the analog for $\mathbf{v}$, which gives

$$\int \dot{\delta}_k e^{i\mathbf{k}\mathbf{x}} d^3k = -\nabla(1 + \int \delta_k e^{i\mathbf{k}\mathbf{x}} d^3k)(\int \mathbf{v}_{k'} e^{i\mathbf{k}'\mathbf{x}} d^3k') \tag{8.23}$$

or

$$\int \dot{\delta}_k e^{i\mathbf{k}\mathbf{x}} d^3k = -i \int \mathbf{v}_k \mathbf{k} e^{i\mathbf{k}\mathbf{x}} d^3k - i \int \delta_k \mathbf{v}_{k'} (\mathbf{k} + \mathbf{k}') e^{i(\mathbf{k}+\mathbf{k}')\mathbf{x}} d^3k d^3k' \tag{8.24}$$

Integrating over $(2\pi)^{-3} e^{-i\mathbf{k}''\mathbf{x}} d^3x$ one gets on the rhs

$$\int \frac{\dot{\delta}_k}{(2\pi)^3} e^{i(\mathbf{k}-\mathbf{k}'')\mathbf{x}} d^3k d^3x = \int \dot{\delta}_k d^3k \delta_D(\mathbf{k} - \mathbf{k}'') = \dot{\delta}_{k''} \tag{8.25}$$

and on the lhs

$$-i\mathbf{v}_{k''}\mathbf{k}'' - i \int \delta_k \mathbf{v}_{k'} (\mathbf{k} + \mathbf{k}') \delta_D(\mathbf{k} + \mathbf{k}' - \mathbf{k}'') d^3k d^3k' \tag{8.26}$$

So we have (switching some k labeling)

$$\dot{\delta}_k + i\mathbf{v}_k\mathbf{k} = -i \int \delta_{k_1} \mathbf{v}_{k_2} (\mathbf{k}_1 + \mathbf{k}_2) \delta_D(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k}) d^3k_1 d^3k_2 \tag{8.27}$$

or also, since $\delta_{k_1} = -i\mathbf{k}_1\mathbf{v}_1/(\mathcal{H}f) = -\theta/(\mathcal{H}f)$ and using Eq. (8.22)

$$\dot{\delta}_k + \theta = \mathcal{H}f \int \delta_{k_1}\delta_{k_2} \frac{\mathbf{k}_2}{k_2^2} \cdot (\mathbf{k}_1 + \mathbf{k}_2) \delta_D(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k}) d^3k_1 d^3k_2 \tag{8.28}$$

We now expand the perturbation variables as

$\delta = G(z)\delta_0$

and the growth rate $f \equiv G'/G$ (we assume G, f are k-independent). Then Eq. (**??**) can be written as (with a new definition of θ: $\theta_{new} = \theta_{old}/\mathcal{H}$)

$$\mathcal{H}\delta' + \mathcal{H}\theta = \mathcal{H}G^2 f \int \delta_1\delta_2 \frac{\mathbf{k}_2\mathbf{k}_1 + \mathbf{k}_2\mathbf{k}_2}{k_2^2} \delta_D(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k}) d^3k_1 d^3k_2 \tag{8.30}$$

where the prime is $d/d\log a$ and where δ_i are evaluated at the present time (so they are independent of time). All δ, θ inside the integrals from now on are the present values. Then we also have

$$\theta + \delta' = G^2 f \int \delta_1\delta_2 K_C \delta_D(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k}) d^3k_1 d^3k_2 \equiv C \tag{8.31}$$

where the kernel K_C can be symmetrized

$$K_C = \frac{1}{2}\left[\frac{\mathbf{k}_2\cdot(\mathbf{k}_1+\mathbf{k}_2)}{k_2^2}+\frac{\mathbf{k}_1\cdot(\mathbf{k}_1+\mathbf{k}_2)}{k_1^2}\right]$$
$$= \frac{1}{2}\left(\frac{\mathbf{k}_2}{k_1^2}+\frac{\mathbf{k}_1}{k_2^2}\right)\cdot(\mathbf{k}_1+\mathbf{k}_2)$$

Similarly, the general Euler equation (8.9) can be written as

$$\theta' + F\theta + S\delta = -\nabla\left[(\int \mathbf{v}_k e^{i\mathbf{k}\mathbf{x}}d^3k)\cdot\nabla\int \mathbf{v}_{k'}e^{i\mathbf{k}'\mathbf{x}}d^3k'\right] \tag{8.32}$$
$$= -i\nabla\left[\int(\mathbf{v}_k\cdot\mathbf{k}')\mathbf{v}_{k'}e^{i\mathbf{k}\mathbf{x}}e^{i\mathbf{k}'\mathbf{x}}d^3kd^3k'\right] \tag{8.33}$$
$$= \int(\mathbf{v}_k\cdot\mathbf{k}')(\mathbf{v}_{k'}\cdot(\mathbf{k}+\mathbf{k}'))e^{i(\mathbf{k}'+\mathbf{k})\mathbf{x}}d^3k'd^3k \tag{8.34}$$

(with $F = 1+\mathcal{H}'/\mathcal{H}, S=-3/2$).

From now on we put ourselves in an Einstein-de Sitter model (i.e. $\Omega_m = 1$). In this case, $G = G' = G'' = a$ and $f = 1$. Moreover, $F = 1/2$. Because of the new definition of θ, we have now at linear level $\delta = -\theta/f = -\theta$, and therefore δ and θ have the same growth factor G. Then we see that

$$\theta' + F\theta + S\delta = -G^2\int\theta_k\theta_{k'}(\frac{\mathbf{k}}{k^2}\cdot\mathbf{k}')(\frac{\mathbf{k}'}{k'^2}\cdot(\mathbf{k}+\mathbf{k}'))e^{i(\mathbf{k}'+\mathbf{k})\mathbf{x}}d^3k'd^3k \tag{8.35}$$
$$= -G^2\int\theta_k\theta_{k'}(\frac{\mathbf{k}\cdot\mathbf{k}'}{k^2k'^2}\mathbf{k}'\cdot(\mathbf{k}+\mathbf{k}'))e^{i(\mathbf{k}'+\mathbf{k})\mathbf{x}}d^3k'd^3k \tag{8.36}$$

The kernel can be symmetrized:

$$K_E = \frac{1}{2k^2k'^2}[\mathbf{k}\cdot\mathbf{k}'(\mathbf{k}'+\mathbf{k})\cdot(\mathbf{k}+\mathbf{k}')] = \frac{\mathbf{k}\cdot\mathbf{k}'(\mathbf{k}'+\mathbf{k})^2}{2k^2k'^2} \tag{8.37}$$

Integrating again over $(2\pi)^{-3}e^{-i\mathbf{k}''\mathbf{x}}d^3x$ one gets

$$\theta' + F\theta + S\delta = -G^2\int\theta_1\theta_2K_E\delta_D(\mathbf{k_1}+\mathbf{k}_2-\mathbf{k})d^3k_1d^3k_2 \tag{8.38}$$

and finally

$$\theta' + F\theta + S\delta = -G^2\int\delta_1\delta_2K_E\delta_D(\mathbf{k}_1+\mathbf{k}_2-\mathbf{k})d^3k_1d^3k_2 \tag{8.39}$$
$$= -G^2\int\theta_1\theta_2K_E\delta_D(\mathbf{k}_1+\mathbf{k}_2-\mathbf{k})d^3k_1d^3k_2 \equiv E \tag{8.40}$$

where (after symmetrization)

$$K_C = \frac{\mathbf{k}}{2}(\frac{\mathbf{k}_1}{k_1^2}+\frac{\mathbf{k}_2}{k_2^2}) \tag{8.41}$$
$$K_E = \frac{\mathbf{k}\cdot\mathbf{k}'(\mathbf{k}'+\mathbf{k})^2}{2k^2k'^2} \tag{8.42}$$

We can now differentiate the θ equation and obtain

$$\theta' + \delta'' = 2G^2\int\delta_1\delta_2K_C\delta_D(\mathbf{k}_1+\mathbf{k}_2-\mathbf{k})d^3k_1d^3k_2 = C' \tag{8.43}$$

Then we insert the θ' equation, replace again $\theta = C-\delta'$ and obtain an equation for the second order δ alone

$$\delta'' + (-F(C-\delta') - S\delta + E) = C'$$

or

$$\delta'' + F\delta' - S\delta = C' - E + FC \tag{8.44}$$

Putting $\delta^{(2)} = G^{(2)}\delta_k^{(2)}$, where $\delta_k^{(2)}$ depends only on $\mathbf{k}$, and, as shown in Eq. (8.20), $G^{(2)} = \alpha G^2$ (with $\alpha = 17/21$), we see that

$$\delta'' + F\delta' - S\delta = [(G^{(2)})'' + FG^{(2)\prime} - SG^{(2)}]\delta = (4 + 2F - S)\alpha G^2\delta_k^{(2)} = \frac{7}{2}\alpha G^2\delta_k^{(2)} = \frac{7}{2}\delta^{(2)} \tag{8.45}$$

then one has

$$\delta^{(2)} = \frac{2}{7}(\delta'' + F\delta' - S\delta) = 2\frac{C' - E + FC}{7} \tag{8.46}$$

$$= \frac{2G^2}{7}\int \delta_1\delta_2[2K_C + K_E + \frac{1}{2}K_C]\delta_D(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k})d^3k_1d^3k_2 \tag{8.47}$$

So we obtain

$$\delta^{(2)} = \frac{G^2}{7}\int \delta_1\delta_2[5K_C + 2K_E]\delta_D(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k})d^3k_1d^3k_2 \tag{8.48}$$

$$= \frac{2G^2}{7}\int \delta_1\delta_2[5\frac{\mathbf{k}}{4}(\frac{\mathbf{k}_1}{k_1^2} + \frac{\mathbf{k}_2}{k_2^2}) + \frac{(\mathbf{k}_1 + \mathbf{k}_2)^2(\mathbf{k}_1\mathbf{k}_2)}{2k_1^2k_2^2}]\delta_D(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k})d^3k_1d^3k_2 \tag{8.49}$$

$$= G^2\int \delta_1\delta_2F_2(\mathbf{k}_1, \mathbf{k}_2)\delta_D(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k})d^3k_1d^3k_2 \tag{8.50}$$

where

$$F_2 = \frac{5K_C + 2K_E}{7} = \frac{2}{7}\left[5\frac{\mathbf{k}}{4}(\frac{\mathbf{k}_1}{k_1^2} + \frac{\mathbf{k}_2}{k_2^2}) + \frac{(\mathbf{k}_1 + \mathbf{k}_2)^2(\mathbf{k}_1\mathbf{k}_2)}{2k_1^2k_2^2}\right] \tag{8.51}$$

$$= \frac{2}{7}\left[\frac{5}{4}(2 + \mathbf{k}_1\mathbf{k}_2(\frac{1}{k_1^2} + \frac{1}{k_2^2})) + \frac{(k_1^2 + k_2^2 + 2\mathbf{k}_1\mathbf{k}_2)(\mathbf{k}_1\mathbf{k}_2)}{2k_1^2k_2^2}\right] \tag{8.52}$$

$$= \frac{2}{7}\left[\frac{5}{2} + \frac{7}{4}\mathbf{k}_1\mathbf{k}_2(\frac{1}{k_1^2} + \frac{1}{k_2^2})) + \frac{(\mathbf{k}_1\mathbf{k}_2)^2}{k_1^2k_2^2}\right] \tag{8.53}$$

$$= \frac{5}{7} + \frac{1}{2}\mathbf{k}_1\mathbf{k}_2(\frac{1}{k_1^2} + \frac{1}{k_2^2})) + \frac{2}{7}\frac{(\mathbf{k}_1\mathbf{k}_2)^2}{k_1^2k_2^2} \tag{8.54}$$

Similarly, one can write an equation for $\theta^{(2)}$ alone

$$\theta^{(2)} + \delta'^{(2)} = \theta^{(2)} + 2GG^{'}\delta^{(2)} = G^2\int \delta_1\delta_2K_C\delta_D(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k})d^3k_1d^3k_2 \tag{8.55}$$

from which, since $GG' = G^2$ in EdS,

$$\theta^{(2)} = -2G^2\int \delta_1\delta_2F_2\delta_D(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k})d^3k_1d^3k_2$$

$$+ G^2\int \delta_1\delta_2K_C\delta_D(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k})d^3k_1d^3k_2$$

or, using again $\theta = -\delta$ for the linear terms inside the integral,

$$\theta^{(2)} = G^2\int \theta_1\theta_2G_2\delta_D(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k})d^3k_1d^3k_2 \tag{8.56}$$

with

$$G_2 = -2F_2 + K_C = -\frac{10}{7} - \mathbf{k}_1\mathbf{k}_2(\frac{1}{k_1^2} + \frac{1}{k_2^2}) - \frac{4}{7}\frac{(\mathbf{k}_1\mathbf{k}_2)^2}{k_1^2k_2^2} + \frac{\mathbf{k}_1 + \mathbf{k}_2}{2}(\frac{\mathbf{k}_1}{k_1^2} + \frac{\mathbf{k}_2}{k_2^2}) \tag{8.57}$$

$$= -\frac{10}{7} - \mathbf{k}_1\mathbf{k}_2(\frac{1}{k_1^2} + \frac{1}{k_2^2}) - \frac{4}{7}\frac{(\mathbf{k}_1\mathbf{k}_2)^2}{k_1^2k_2^2} + 1 + \frac{1}{2}(\frac{\mathbf{k}_2\mathbf{k}_1}{k_1^2} + \frac{\mathbf{k}_1\mathbf{k}_2}{k_2^2}) \tag{8.58}$$

$$= -\frac{3}{7} - \mathbf{k}_1\mathbf{k}_2(\frac{1}{k_1^2} + \frac{1}{k_2^2}) - \frac{4}{7}\frac{(\mathbf{k}_1\mathbf{k}_2)^2}{k_1^2k_2^2} + \frac{1}{2}\mathbf{k}_2\mathbf{k}_1(\frac{1}{k_1^2} + \frac{1}{k_2^2}) \tag{8.59}$$

$$= -\frac{3}{7} - \frac{1}{2}\mathbf{k}_2\mathbf{k}_1(\frac{1}{k_1^2} + \frac{1}{k_2^2}) - \frac{4}{7}\frac{(\mathbf{k}_1\mathbf{k}_2)^2}{k_1^2k_2^2} \tag{8.60}$$

In general, it happens that $G^{(2)} \propto G^2$ even for non EdS models.

Whatever the cosmology, therefore, one always has the forms

$$\theta^{(2)} = G^2 \int \theta_1\theta_2 G_2(\mathbf{k}_1, \mathbf{k}_2)\delta_D(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k})d^3k_1d^3k_2 \tag{8.61}$$

$$\delta^{(2)} = G^2 \int \delta_1\delta_2 F_2(\mathbf{k}_1, \mathbf{k}_2)\delta_D(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k})d^3k_1d^3k_2 \tag{8.62}$$

where G is the linear growth function and

$$F_2 = \frac{5}{7} + \frac{\mathbf{k}_1\mathbf{k}_2}{2k_1k_2}(\frac{k_1}{k_2} + \frac{k_2}{k_1}) + \frac{2}{7}(\frac{\mathbf{k}_1\mathbf{k}_2}{k_1k_2})^2 \tag{8.63}$$

$$G_2 = -\left[\frac{3}{7} + \frac{\mathbf{k}_1\mathbf{k}_2}{2k_1k_2}(\frac{k_1}{k_2} + \frac{k_2}{k_1}) + \frac{4}{7}(\frac{\mathbf{k}_1\mathbf{k}_2}{k_1k_2})^2\right] \tag{8.64}$$

Note that

$$F_2(\mathbf{k}_1, -\mathbf{k}_1) = 0 \tag{8.65}$$

$$G_2(\mathbf{k}_1, -\mathbf{k}_1) = 0 \tag{8.66}$$

The procedure can be extended to all orders by an iterative process. So one gets for instance

$$\theta^{(3)} = G^3 \int \theta_1\theta_2\theta_3 G_3(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)\delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 - \mathbf{k})d^3k_1d^3k_2d^3k_3 \tag{8.67}$$

$$\delta^{(3)} = G^3 \int \delta_1\delta_2\delta_3 F_3(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)\delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 - \mathbf{k})d^3k_1d^3k_2d^3k_3 \tag{8.68}$$

As will be shown shortly, we need also the third order for δ to derive the corrected power spectrum.

8.3 Bias and RSD

So far we have calculated the perturbations of matter at second order. What we observe, though, are galaxies, i.e. a biased traced of the underlying matter perturbations. Moreover, the apparent distribution of galaxies is distorted because we observe in redshift space, not in real space. As can be seen in *Lecture notes on Cosmology*, these two effect modify the linear power spectrum as follows

$$P_{gg,L} = \langle\delta_g^{(1)}\delta_g^{(1)*}\rangle = b^2(1 + \beta\mu^2)^2G^2P_L \tag{8.69}$$

where b is the linear bias (defined as $\delta_g = b\delta_m$) and $\beta = f/b$ is the redshift distortion (RSD) parameter. We need now to see how bias and RSD modify the non-linear spectrum.

At higher order, we need first to generalize the bias. Various tentative schemes have been introduced in the literature, since we know very little about the exact mechanism of bias. A very general form is given e.g. in Chudaykin et al (2020)

$$\delta_g = b\delta_m + \frac{b_2^2}{2}\delta_m^2 + b_G\mathcal{G}_2 + b_\Gamma\Gamma_3 + ... \tag{8.70}$$

where

$$\mathcal{G}_2 = (\partial_i\partial_j\Phi_g)^2 - (\partial_i^2\Phi_g)^2 \tag{8.71}$$
$$\Gamma_3 = \mathcal{G}_2(\Phi_g) - \mathcal{G}_2(\Phi_v) \tag{8.72}$$

and where Φ_g is the gravitational potential and Φ_v the velocity potential. One should then convert the second-order calculation from δ_m and $\Phi_{g,v}$ to δ_g, and convert the real space coordinates into the redshift space ones, resulting in modified kernels. Now there are four bias parameters, $b = b_1, b_2, b_G, b_\Gamma$ that are only time-dependent, and the Z_2, Z_3 functions that replace F_2, F_3 are

$$\begin{aligned} Z_2(\mathbf{k}_a,\mathbf{k}_b) &= b\{F_2(\mathbf{k}_a,\mathbf{k}_b) + \beta\mu^2 G_2(\mathbf{k}_a,\mathbf{k}_b) \\ &+ \frac{\beta b\mu k}{2}\left[\frac{\mu_{az}}{k_a}(1+\beta_b\mu_{bz}^2) + \frac{\mu_{bz}}{k_b}(1+\beta_a\mu_{az}^2)\right]\} + \frac{b_2}{2} + b_G S_1(\mathbf{k}_a,\mathbf{k}_b) \end{aligned} \tag{8.73}$$

(already symmetrized) and

$$\begin{aligned} Z_3(\mathbf{k}_a,\mathbf{k}_b,\mathbf{k}_c) &= b\{F_3(\mathbf{k}_a,\mathbf{k}_b,\mathbf{k}_c) + \beta\mu^2 G_3(\mathbf{k}_a,\mathbf{k}_b,\mathbf{k}_c) + \beta\mu kb[F_2(\mathbf{k}_a,\mathbf{k}_b) + \beta_{ab}\mu_{abz}^2 G_2(\mathbf{k}_a,\mathbf{k}_b)]\frac{\mu_{cz}}{k_c} \\ &+ \beta\mu kb(1+\beta_a\mu_{az}^2)\frac{\mu_{bcz}}{k_{bc}}G_2(\mathbf{k}_b,\mathbf{k}_c) + \frac{(\beta\mu k)^2}{2}b^2(1+\beta_a\mu_{az}^2)\frac{\mu_{bz}}{k_b}\frac{\mu_{cz}}{k_c}\} \\ &+ 2b_G S_1(\mathbf{k}_a,\mathbf{k}_b+\mathbf{k}_c)F_2(\mathbf{k}_b,\mathbf{k}_c) + b_G b\beta\mu k\frac{\mu_{az}}{k_a}S_1(\mathbf{k}_b,\mathbf{k}_c) \\ &+ 2b_\Gamma S_1(\mathbf{k}_a,\mathbf{k}_b+\mathbf{k}_c)(F_2(\mathbf{k}_b,\mathbf{k}_c) - G_2(\mathbf{k}_b,\mathbf{k}_c)) \end{aligned} \tag{8.74}$$

(to be symmetrized), where terms in b_2 have been discarded because degenerate with other terms (also a b_3 term which should appear here has been discarded for the same reason), and

$$S_1(\mathbf{k}_a,\mathbf{k}_b) = \frac{(\mathbf{k}_a\cdot\mathbf{k}_b)^2}{k_a^2k_b^2} - 1 \tag{8.75}$$

Indexes a, b, etc refer to the $\mathbf{k}$ vectors: e.g. $\beta_a = \beta(k_a)$. Double letters, e.g. ab, refer to $\mathbf{k}_a + \mathbf{k}_b$. All μ's are the angles wrt the line of sight $\hat{z}$, except μ_1 , which is the angle between k and k_1. All angles can be expressed in terms of μ, μ_1, ϕ_1 by the relations

$$\mu_{bz} = \frac{\mathbf{k}_b\mathbf{z}}{k_b} = \frac{(\mathbf{k}-\mathbf{k}_1)\mathbf{z}}{|\mathbf{k}-\mathbf{k}_1|} = \frac{-k\mu + k_1\mu\mu_1 + k_1(1-\mu^2)^{1/2}(1-\mu_1^2)^{1/2}\cos\phi_1}{(k_1^2+k^2-2kk_1\mu_1)^{1/2}} \tag{8.76}$$
$$\mu_{az} = \frac{\mathbf{k}_a\mathbf{z}}{k_1} = \mu\mu_1 + (1-\mu^2)^{1/2}(1-\mu_1^2)^{1/2}\cos\phi_1 \tag{8.77}$$

and μ_{abz}, μ_{bcz} denote the angles between $\mathbf{k}_a + \mathbf{k}_b$ and $\hat{z}$ etc. The ϕ_1 integration is always analytical and can be included in the definition of the kernels.

8.4 Power spectrum at one loop

As we have already mentioned, the linear power spectrum including bias and redshift distortion is

$$P_{gg,L} = \langle\delta_g^{(1)}\delta_g^{(1)*}\rangle = b^2(1+\beta\mu^2)^2G^2P_L \tag{8.78}$$

where P_L is the matter power spectrum at $z = 0$. Analogously, the power spectrum up to fourth order (the third cancels out because the linear δ are Gaussian) is then

$$\begin{aligned} P_{gg} &= \langle\delta_g\delta_g\rangle = \langle(\delta^{(1)}+\delta^{(2)}+\delta^{(3)})_g(\delta^{(1)*}\delta^{(2)*}\delta^{(3)*})_g\rangle \\ &= (b+f\mu^2)^2G^2P_L + G^4\int\langle\delta_{k_1}^*\delta_{k_2}^*\delta_{k_3}\delta_{k_4}\rangle Z_2(k_1,k_2)Z_2(k_3,k_4)\delta_D(\mathbf{k}_1+\mathbf{k}_2-\mathbf{k})\delta_D(\mathbf{k}_3+\mathbf{k}_4-\mathbf{k})d^3k_1d^3k_2d^3k_3d^3k_4 \\ &+ 2(b+f\mu^2)G^4\int\langle\delta_k^*\delta_{k_1}\delta_{k_2}\delta_{k_3}\rangle Z_3(k_1,k_2,k_3)\delta_D(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3-\mathbf{k})d^3k_1d^3k_2d^3k_3 \end{aligned} \tag{8.79}$$

(all power spectra are taken at the present time). Now assuming the linear δ are Gaussian variable, one has due to Wick's theorem that

$$\langle\delta^*_{k_1}\delta^*_{k_2}\delta_{k_3}\delta_{k_4}\rangle = \langle\delta^*_{k_1}\delta^*_{k_2}\rangle\langle\delta_{k_3}\delta_{k_4}\rangle + \langle\delta^*_{k_1}\delta_{k_3}\rangle\langle\delta^*_{k_2}\delta_{k_4}\rangle + \langle\delta^*_{k_1}\delta_{k_4}\rangle\langle\delta^*_{k_2}\delta_{k_3}\rangle \tag{8.80}$$

where

$$\langle\delta_{k_1}\delta^*_{k_2}\rangle = \langle\delta_{k_1}\delta_{-k_2}\rangle = P(k_1)\delta_D(\mathbf{k}_1 - \mathbf{k}_2) \tag{8.81}$$

So we have, e.g. for the first term, $\langle\delta^*_{k_1}\delta^*_{k_2}\rangle = P(k_1)\delta_D(\mathbf{k}_1 + \mathbf{k}_2)$ and therefore

$$\int \langle\delta^*_{k_1}\delta^*_{k_2}\rangle\langle\delta_{k_3}\delta_{k_4}\rangle Z_2(\mathbf{k}_1,\mathbf{k}_2)Z_2(\mathbf{k}_3,\mathbf{k}_4)\delta_D(\mathbf{k}_1+\mathbf{k}_2-\mathbf{k})\delta_D(\mathbf{k}_3+\mathbf{k}_4-\mathbf{k})d^3k_1d^3k_2d^3k_3d^3k_4 =$$
$$\int P(k_1)\delta_D(\mathbf{k}_1+\mathbf{k}_2)P(k_3)\delta_D(\mathbf{k}_3+\mathbf{k}_4)Z_2(\mathbf{k}_1,\mathbf{k}_2)Z_2(\mathbf{k}_3,\mathbf{k}_4)\delta_D(\mathbf{k}_1+\mathbf{k}_2-\mathbf{k})\delta_D(\mathbf{k}_3+\mathbf{k}_4-\mathbf{k})d^3k_1d^3k_2d^3k_3d^3k_4 =$$
$$\int P(k_1)P(k_3)Z_2(\mathbf{k}_1,-\mathbf{k}_1)Z_2(\mathbf{k}_3,-\mathbf{k}_3)\delta_D(-\mathbf{k})\delta_D(-\mathbf{k})d^3k_1d^3k_3 = 0 \tag{8.82}$$

The other two terms instead do not vanish and give two identical contributions of this form:

$$\int P(k_1)P(k_4)Z_2(k_1,k_4)Z_2(k_1,k_4)\delta_D(\mathbf{k}_1+\mathbf{k}_4-\mathbf{k})\delta_D(\mathbf{k}_1+\mathbf{k}_4-\mathbf{k})d^3k_1d^3k_4 =$$
$$\int P(k_1)P(|\mathbf{k}-\mathbf{k}_1|)Z_2^2(\mathbf{k}_1,\mathbf{k}-\mathbf{k}_1)d^3k_1 \equiv P_{22} \tag{8.83}$$

Similarly,

$$\langle\delta^*_k\delta_{k_1}\delta_{k_2}\delta_{k_3}\rangle = \langle\delta^*_k\delta_{k_1}\rangle\langle\delta_{k_2}\delta^*_{-k_3}\rangle\delta_D(k-k_1)\delta_D(k_2+k_3) + (1\leftrightarrow 2) + (2\leftrightarrow 3) \tag{8.84}$$

and each of the three term equals

$$\int \langle\delta^*_k\delta_{k_1}\rangle\langle\delta_{k_2}\delta^*_{-k_3}\rangle\delta_D(\mathbf{k}-\mathbf{k}_1)\delta_D(\mathbf{k}_2+\mathbf{k}_3)Z_3(k_1,k_2,k_3)\delta_D(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3-\mathbf{k})d^3k_1d^3k_2d^3k_3 =$$
$$\int P(k_1)P(k_2)\delta_D(\mathbf{k}_2+\mathbf{k}_3)Z_3(\mathbf{k},\mathbf{k}_2,\mathbf{k}_3)\delta_D(\mathbf{k}_2+\mathbf{k}_3)d^3k_2d^3k_3 =$$
$$\int P(k)P(k_1)Z_3(\mathbf{k},\mathbf{k}_1,-\mathbf{k}_1)d^3k_1 \equiv P_{31} \tag{8.85}$$

(changing the subscript of k in the last step). So finally

$$P_{gg}(\mathbf{k}) = (b+f\mu^2)^2G^2P_L + 2G^4P_{22} + 6(b+f\mu^2)G^4P_{31} \tag{8.86}$$

This non-linear power spectrum, also called one-loop $P(k)$, is however still not a good approximation to N-body simulations even for relatively small $k < 0.2h$/Mpc, mostly due to the neglect of rotational velocity components. To improve the agreement one has to add semi-empirical terms, called counterterms, of the form (see Fig. 8.1)

$$P_{ctr} = -2P(k)k^2(c_0 + c_2\beta\mu^2 + c_4\beta^2\mu^4) \tag{8.87}$$

So finally

$$P_{gg}(k,\mu,z) = B^2P_L + 2P_{22} + 6BP_{31} - 2P(k)k^2(c_0 + c_2\beta\mu^2 + c_4\beta^2\mu^4) \tag{8.88}$$

where $B = b(1+\beta\mu^2)$.

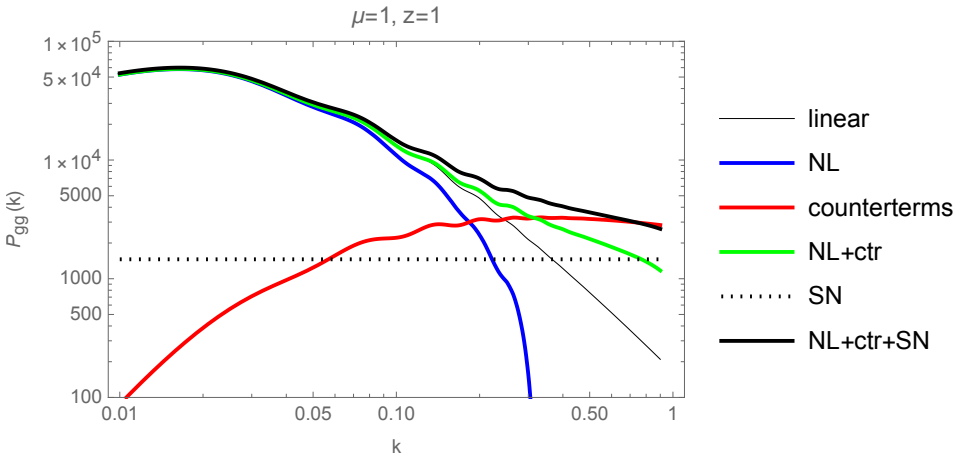

Figure 8.1: Linear and non-linear spectra at $z = 1$.

Part II

Observations

Chapter 9

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.

9.1 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 $m_{\rm th} = 5\log_{10} d_L(z;\Omega_m^{(0)},\Omega_\Lambda^{(0)}) + \text{constant}$, where d_L is the luminosity distance.

If we repeat the measure and we obtain $x_1, x_2, x_3...$, then the law of joint probability tells us that the probability of obtaining x_1 in the interval $\mathrm{d}x_1$ around x_1, x_2 in the interval $\mathrm{d}x_2$ around x_2 and so forth is

$$f(x_i;\theta)\mathrm{d}^n x_i \equiv \prod_i f_i(x_i;\theta)\mathrm{d}x_i = f_1(x_1;\theta)f_2(x_2;\theta)f_3(x_3;\theta)...\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3...\,, \tag{9.1}$$

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 that value maximizing $f(x;\theta)$. 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 maximizes the joint function $f(x_1, x_2, ..x_n;\theta_1,\theta_2,...\theta_m)$. 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

$$\frac{\partial f(x_i;\theta_j)}{\partial\theta_j} = 0\,, \qquad j = 1,..,m\,. \tag{9.2}$$

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_is; 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' theorem[a]:

$$P(T;D) = \frac{P(D;T)P(T)}{P(D)}\,, \tag{9.3}$$

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 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

$$L(\theta_j;x_i) = \frac{f(x_i;\theta_j)p(\theta_j)}{g(x_i)}\,, \tag{9.4}$$

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:

$$\int L(\theta_j;x_i)\mathrm{d}^n\theta_j = 1 = \frac{\int f(x_i;\theta_j)p(\theta_j)\mathrm{d}^n\theta_j}{g(x_i)}\,, \tag{9.5}$$

and therefore

$$\int f(x_i;\theta_j)p(\theta_j)\mathrm{d}^n\theta_j = g(x_i)\,. \tag{9.6}$$

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 ΛCDM 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 0.78 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

[a]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.

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 advantage 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 (9.2) is then replaced by

$$\frac{\partial L(\theta_i)}{\partial \theta_i} = 0\,, \qquad i = 1, .., n\,. \tag{9.7}$$

If, as usually the case, we discard the denominator $g(x_i)$ in Eq. (9.4), the posterior L is not normalized and its normalization has to be recalculated. The overall normalization N is the integral over the parameter space:

$$N = \int L(\theta_i)\, \mathrm{d}^n\theta_i\,, \tag{9.8}$$

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 $R(\alpha)$ of constant $L(\theta_i)$ for which

$$\int_{R(\alpha)} L(\theta_i)\, \mathrm{d}^n\theta = \alpha\,. \tag{9.9}$$

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

$$\int \hat{L}(L_i)\, \mathrm{d}^n\theta = \alpha_i\,, \tag{9.10}$$

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" of 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 $m_{\rm th} = 5 \log_{10} d_L(z; \Omega_m^{(0)}, \Omega_\Lambda^{(0)}) + C$, we may be interested in $\Omega_m^{(0)}, \Omega_\Lambda^{(0)}$ but not in the constant C. 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:

$$L(\Omega_m^{(0)}, \Omega_\Lambda^{(0)}) \equiv \int L(C, \Omega_m^{(0)}, \Omega_\Lambda^{(0)})\, \mathrm{d}C\,, \tag{9.11}$$

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. (9.11)]. This very common procedure is called *marginalization*.

Often one wants to marginalize a multidimensional problem down to a more manageable and plottable 2-dimensional likelihood. Also, one could quote final confidence regions by marginalizing in turn to single

parameters, e.g.,

$$L(\Omega_\Lambda^{(0)}) = \int_0^\infty L(\Omega_m^{(0)}, \Omega_\Lambda^{(0)})\, \mathrm{d}\Omega_m^{(0)}\,. \tag{9.12}$$

For instance, if the maximum likelihood estimator of $\Omega_m^{(0)}$ is 0.3 and

$$\int_R L(\Omega_m^{(0)})\mathrm{d}\Omega_m^{(0)} = 0.683\,, \tag{9.13}$$

when R is the interval $\Omega_m^{(0)} = [0.1, 0.4]$, we will write as our final result $\Omega_m^{(0)} = 0.3^{+0.1}_{-0.2}$ 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 the 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 $m_{\mathrm{th},i} = M + 5\log_{10} d_L(z_i;\theta_j) + 25$, 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 $m_{\mathrm{th},i} = \alpha + \mu_i$, 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:

$$L(\theta_j) = N\int \mathrm{d}\alpha\, \exp\left[-\frac{1}{2}\sum_i \frac{(m_i - \mu_i - \alpha)^2}{\sigma_i^2}\right]\,, \tag{9.14}$$

where N is an unimportant normalization factor. Then we have

$$\begin{aligned} L(\theta_j) &= N\int \mathrm{d}\alpha \exp\left[-\frac{1}{2}\sum_i \frac{(m_i-\mu_i)^2 + \alpha^2 - 2\alpha(m_i-\mu_i)}{\sigma_i^2}\right] \\ &= N\exp(-S_2/2)\int \mathrm{d}\alpha \exp(\alpha S_1 - \alpha^2 S_0/2) \\ &= N\exp\left[-\frac{1}{2}\left(S_2 - \frac{S_1^2}{S_0}\right)\right]\int \mathrm{d}\alpha\, \exp\left[-\frac{1}{2}\left(\alpha - \frac{S_1}{S_0}\right)^2 S_0\right]\,, \end{aligned} \tag{9.15}$$

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:

$$L(\theta_j) = N\exp\left[-\frac{1}{2}\left(S_2 - \frac{S_1^2}{S_0}\right)\right]\,. \tag{9.16}$$

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 $f(x_i;\mu,\sigma^2)$

$$f(x_i;\mu,\sigma^2) = (2\pi\sigma^2)^{-n/2}\exp\left[-\frac{1}{2}\sum_i^n \frac{(x_i-\mu)^2}{\sigma^2}\right]\,, \tag{9.17}$$

then the maximum likelihood estimator of μ is given by

$$\hat{\mu} = \frac{1}{n}\sum_{i}^{n} x_i \,. \tag{9.18}$$

Analogously, you can prove that the maximum likelihood variance estimator is

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i}^{n}(x_i - \hat{\mu})^2 \,. \tag{9.19}$$

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 $n \to \infty$) 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.

9.2 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 $w_{\rm DE} = \text{constant}$ 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 $w_{\rm DE} = \text{constant}$?

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. (9.6) that the evidence is defined as the likelihood integral over the whole domain

$$E(\boldsymbol{x}; M) = \int f(\boldsymbol{x}; \theta_i^M) p(\theta_i^M)\, \mathrm{d}^n\theta_i^M \,, \tag{9.20}$$

where as before $\boldsymbol{x} = (x_1, x_2, ...)$ 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 [542]

$$B_{12} = \frac{\int f(\boldsymbol{x}; \theta_i^{M_1}) p(\theta_i^{M_1}) \mathrm{d}^n\theta_i^{M_1}}{\int f(\boldsymbol{x}; \theta_i^{M_2}) p(\theta_i^{M_2}) \mathrm{d}^n\theta_i^{M_2}} \,. \tag{9.21}$$

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

$$L(M; \boldsymbol{x}) \propto E(\boldsymbol{x}; M) p(M) \,, \tag{9.22}$$

and evaluate the ratio

$$\frac{L(M_1; \boldsymbol{x})}{L(M_2; \boldsymbol{x})} = B_{12} \frac{p(M_1)}{p(M_2)} \,. \tag{9.23}$$

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 $p(\theta_i) = \Pi_i p_i(\theta_i)$, we see that the integral over θ_n decouples. Since the priors are just standard probability distribution functions we have $\int p_n(\theta_n) \mathrm{d}\theta_n = 1$, 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

$$\begin{aligned} L(\theta_i) &= \prod_i f(\boldsymbol{x}; \theta_i) p(\theta_i) \\ &= \prod_i f_{\max,i} (2\pi\sigma_{P,i}^2)^{-1/2} \exp\left[-\frac{(\theta_i - \theta_i^{(B)})^2}{2\sigma_{B,i}^2} - \frac{(\theta_i - \theta_i^{(P)})^2}{2\sigma_{P,i}^2} \right] \\ &= \prod_i f_{\max,i} (2\pi\sigma_{P,i}^2)^{-1/2} \exp\left[-\frac{1}{2} \frac{(\theta_i - \theta_i^*)^2}{\sigma_{i*}^2} \right] \exp\left[-\frac{1}{2} \frac{(\theta_i^{(B)} - \theta_i^{(P)})^2}{\sigma_{B,i}^2 + \sigma_{P,i}^2} \right] \,, \end{aligned} \tag{9.24}$$

where $f_{\max,i}$ is the i-th likelihood maximum and where the posterior mean and variance for each i are

$$\theta_i^* = \frac{\sigma_{B,i}^2 \theta_i^{(P)} + \sigma_{P,i}^2 \theta_i^{(B)}}{\sigma_{B,i}^2 + \sigma_{P,i}^2} \,, \tag{9.25}$$

$$\sigma_{i*}^2 = \frac{\sigma_{P,i}^2 \sigma_{B,i}^2}{\sigma_{B,i}^2 + \sigma_{P,i}^2} \,. \tag{9.26}$$

The evidence is therefore

$$\begin{aligned} E &= \int f(\boldsymbol{x}; \theta_i) p(\theta_i) \mathrm{d}\theta_i \\ &= \prod_i f_{\max,i} \frac{\sigma_{i*}}{\sigma_{P,i} \exp\left\{ -\frac{1}{2} \left[\left(\frac{\theta_i^{(B)}}{\sigma_{B,i}} \right)^2 + \left(\frac{\theta_i^{(P)}}{\sigma_{P,i}} \right)^2 - \left(\frac{\theta_i^*}{\sigma_{i*}} \right)^2 \right] \right\}} \,. \end{aligned} \tag{9.27}$$

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 Occam's razor argument[b]. 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 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 the problem [13.3].

It is a matter of straightforward algebra to extend the expression to correlated Gaussian parameters. If the evidence integral is

$$\begin{aligned} E &= \int f(\boldsymbol{x};\theta_i)p(\theta_i)\mathrm{d}\theta_i \\ &\approx f_{\max}\int \exp\left[-\frac{1}{2}(\theta_i-\theta_i^{(B)})L_{ij}(\theta_j-\theta_j^{(B)}) - \frac{1}{2}(\theta_i-\theta_i^{(P)})P_{ij}(\theta_j-\theta_j^{(P)})\right]\mathrm{d}\theta_i\,, \end{aligned} \tag{9.28}$$

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

$$E = f_{\max}\frac{|\boldsymbol{F}|^{-1/2}}{|\boldsymbol{P}|^{-1/2}}\exp\left[-\frac{1}{2}(\theta_i^{(B)}L_{ij}\theta_j^{(B)} + \theta_i^{(P)}P_{ij}\theta_j^{(P)} - \tilde{\theta}_i F_{ij}\tilde{\theta}_j)\right]\,, \tag{9.29}$$

where $\boldsymbol{F} = \boldsymbol{P} + \boldsymbol{L}$ and $\tilde{\theta}_i = (\boldsymbol{F}^{-1})_{im}[L_{mj}\theta_j^{(B)} + P_{mj}\theta_j^{(P)}]$.

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 [543]). They are however only justified in specific cases and may give conflicting results, sometimes leading to controversies [544, 545]. 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.

9.3 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, ten values for each dimension in parameters space. If there are 10 parameters, this

[b] 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 apocriphal) - one should not multiply entities beyond necessity.

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 a 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,

$$L \approx N \exp\left[-\frac{1}{2}(\theta_i - \hat{\theta}_i)F_{ij}(\theta_j - \hat{\theta}_j)\right], \tag{9.30}$$

where the values $\hat{\theta}_i$, the maximum likelihood estimators, are function 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 (9.30) 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

$$\ln L(\theta_i) \approx \ln L(\hat{\theta}_i) + \frac{1}{2}\left.\frac{\partial^2 \ln L(\theta_i)}{\partial\theta_i \partial\theta_j}\right|_{\rm ML} (\theta_i - \hat{\theta}_i)(\theta_j - \hat{\theta}_j)\,, \tag{9.31}$$

(naturally the first derivatives are absent because they vanish at the peak) we find, comparing with Eq. (9.30), that the normalization $N = L(\hat{\theta}_i)$ depends only on the data and that the *Fisher matrix* (FM) is defined as

$$F_{ij} \equiv -\left.\frac{\partial^2 \ln L(\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}\right|_{\rm ML}. \tag{9.32}$$

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.

$$F_{ij} \equiv -\left\langle \frac{\partial^2 \ln L(\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j} \right\rangle = -\int \frac{\partial^2 \ln L(\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j} L(\boldsymbol{x};\boldsymbol{\theta}) {\rm d}\boldsymbol{x}\,. \tag{9.33}$$

However, within the approximation (9.30), 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 multi-dimensional 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

$$\mathcal{L}_{,\theta}(\theta) \approx \mathcal{L}_{,\theta}(\theta^{(0)}) + \mathcal{L}_{,\theta\theta}(\theta - \theta^{(0)})\,, \tag{9.34}$$

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

$$\theta^{(1)} = \theta^{(0)} - \left.\frac{\mathcal{L}_{,\theta}}{\mathcal{L}_{,\theta\theta}}\right|_{\theta^{(0)}}, \tag{9.35}$$

which could be iterated by assuming as 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 multi-dimensional 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 $m_{\rm th}(z_i; \Omega_m^{(0)}, \Omega_\Lambda^{(0)}) = 5\log_{10} d_L(z_i; \Omega_m^{(0)}, \Omega_\Lambda^{(0)}) + C$. Here we take the PDF to be Gaussian but we could also assume any other PDF that we have any reason to describe the data. Since the data PDF is assumed to be Gaussian we can immediately form the likelihood (neglecting the normalization constant):

$$L_m \approx \exp\left[-\frac{1}{2}\sum_i \frac{(m_i - m_{\rm th}(z_i))^2}{\sigma_i^2}\right] = \exp\left(-\frac{1}{2}\mu_i C_{ij}^{-1}\mu_j\right) . \tag{9.36}$$

Here we have expressed the argument of the exponential in a slightly more general way: we have introduced the vector $\mu_i \equiv m_i - m_{\rm th}(z_i)$ and the correlation matrix C_{ij}, that in this particular case is rather trivial

$$\boldsymbol{C} = {\rm diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2 ...) . \tag{9.37}$$

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. (9.30) like

$$L(\Omega_m^{(0)}, \Omega_\Lambda^{(0)}) = \exp\left[-\frac{1}{2}(\Omega_i^{(0)} - \hat{\Omega}_i^{(0)}) F_{ij} (\Omega_j^{(0)} - \hat{\Omega}_j^{(0)})\right] , \tag{9.38}$$

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 $m_{\rm th}(z_i; \Omega_m^{(0)F}, \Omega_\Lambda^{(0)F})$ 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 (9.36) is then

$$F_{ij} = -\frac{\partial \ln L_m}{\partial\Omega_i^{(0)}\partial\Omega_j^{(0)}}\bigg|_F = \sum_n \frac{1}{\sigma_i^2}\frac{\partial^2 m_{\rm th}(z_n; \Omega_m^{(0)}, \Omega_\Lambda^{(0)})}{\partial\Omega_i^{(0)}\partial\Omega_j^{(0)}}\bigg|_F . \tag{9.39}$$

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 $m_{\rm th}(z_n)$, we vary the likelihood of obtaining *all* m_i by varying $\Omega_{m,\Lambda}^{(0)}$. We can now use Eq. (9.38) 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 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 approximated likelihood is

$$L = \exp\left(-\frac{1}{2}\tilde{x}_i F_{ij}^{(x)} \tilde{x}_j\right) , \tag{9.40}$$

where $\tilde{x}_i = x_i - x_i^{\rm ML}$. Approximating y_j near $x_i^{\rm ML}$ as

$$y_j \approx y_j^{\rm ML} + \left.\frac{\partial y_j}{\partial x_i}\right|_{\rm ML} (x_i - x_i^{\rm ML})\,, \tag{9.41}$$

where $y_j^{\rm ML} \equiv y_j(x^{\rm ML})$, we can write

$$\tilde{y}_j \equiv y_j - y_j^{\rm ML} = J_{ji}^{-1}\tilde{x}_i\,. \tag{9.42}$$

Here $J_{ji} \equiv (\partial x_j/\partial y_i)_{\rm ML}$ is the transformation Jacobian evaluated on the ML estimators. Then we have

$$\tilde{x}_i = J_{i\ell}\tilde{y}_\ell\,, \tag{9.43}$$

and we can find the new Fisher matrix by substituting into Eq. (9.40) simply as

$$F_{ij}^{(y)} = J_{i\ell}F_{\ell m}^{(x)}J_{mj}\,, \tag{9.44}$$

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 $\mathrm{d}x_i\mathrm{d}x_2...$ to the new element $\mathrm{d}y_i\mathrm{d}y_2...$, 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 $|J_{\rm ML}|$ 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 $w_{\rm DE}(a) = w_0 + w_1(1-a)$ [Eq. (1.97)]. We could however have expanded $w_{\rm DE}(a)$ around any other point a_p and write instead $w_{\rm DE}(a) = w_p + w_1(a_p - a)$, where

$$w_p = w_0 + w_1(1 - a_p)\,. \tag{9.45}$$

We can now ask the question whether the constraint we obtain on w_p (i.e. $\sigma^2_{w_0}$) is tighter than the one on w_0, that is whether we can better rule out say $w_{\rm DE} = -1$ 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 $w_{\rm DE}(a)$. 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,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}a}\left[\langle[(w_0 - \hat{w}_0) + (1-a)(w_1 - \hat{w}_1)]^2\rangle\right] &= \frac{\mathrm{d}}{\mathrm{d}a}\left[\sigma^2_{w_0} + (1-a)^2\sigma^2_{w_1} + 2(1-a)\rho\sigma_{w_0}\sigma_{w_1}\right] \\ &= -2(1-a)\sigma^2_{w_1} - 2\rho\sigma_{w_0}\sigma_{w_1} = 0\,. \end{aligned}\tag{9.46}$$

Here $\sigma^2_{w_i} \equiv \langle(w_i - \hat{w}_i)^2\rangle$ 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 [546]

$$a_p = 1 + \frac{\rho\sigma_{w_0}}{\sigma_{w_1}}\,. \tag{9.47}$$

In terms of the two-dimensional Fisher matrix F_{ij} for w_0, w_1, we can write

$$\sigma^2_{w_0} = (\boldsymbol{F}^{-1})_{11}\,, \qquad \sigma^2_{w_1} = (\boldsymbol{F}^{-1})_{22}\,, \qquad \rho\sigma_{w_0}\sigma_{w_1} = (\boldsymbol{F}^{-1})_{12}\,. \tag{9.48}$$

The transformation from $\boldsymbol{p} = (w_0, w_1)$ to $\boldsymbol{q} = (w_p, w_1)$ is achieved by using Eq. (9.44) with the transformation matrix

$$\boldsymbol{J} = \frac{\partial \boldsymbol{p}}{\partial \boldsymbol{q}} = \begin{pmatrix} 1 & 1 - a_p \\ 0 & 1 \end{pmatrix}. \tag{9.49}$$

It is straightforward to verify that with this transformation the new matrix $\boldsymbol{F}_p = \boldsymbol{J}^t\boldsymbol{F}\boldsymbol{J}$ is diagonal (the superscript t denotes transpose) and its inverse is:

$$\boldsymbol{F}_p^{-1} = \begin{pmatrix} \sigma^2_{w_0}(1-\rho^2) & 0 \\ 0 & \sigma^2_{w_1} \end{pmatrix}. \tag{9.50}$$

The parameters w_p, w_1 are therefore uncorrelated and their confidence regions are circular. Moreover, as expected, the error on w_p, $\sigma^2_{w_p} \equiv \sigma^2_{w_0}(1-\rho^2)$, is always smaller than $\sigma^2_{w_0}$.

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 2-dimensional Gaussian PDF

$$G(x_1, x_2) = N \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{x_1^2}{\sigma_1^2} + \frac{x_2^2}{\sigma_2^2} - 2\frac{\rho x_1 x_2}{\sigma_1 \sigma_2}\right)\right], \tag{9.51}$$

where ρ is the correlation factor. This PDF can be written as

$$G(X_i) = N \exp\left[-\frac{1}{2}(X_i C_{ij}^{-1} X_j)\right], \tag{9.52}$$

where $X_i \equiv x_i - \mu_i$ (generalizing to non-zero μ's), and

$$\boldsymbol{C} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}. \tag{9.53}$$

Let us now evaluate the integral $\int G(x_1, x_2)\, \mathrm{d}x_2$ over the whole real domain. The result is given by

$$G(x_1) = \tilde{N} \exp[-x_1^2/(2\sigma_1^2)], \tag{9.54}$$

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 $\boldsymbol{F} = \boldsymbol{C}^{-1}$ we see that the outcome of the marginalization has been the removal from $\boldsymbol{F}^{-1} = \boldsymbol{C}$ 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 $\boldsymbol{F}^{-1}$ 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)

$$\sigma_i^2 = (\boldsymbol{F}^{-1})_{ii}. \tag{9.55}$$

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 [546].

Often we want to reduce the Fisher matrix to a 2×2 matrix $\boldsymbol{F}_2$ for two parameters, say θ_1, θ_2, because then it is easy to plot the resulting 2-dimensional confidence regions, defined as the regions of constant likelihood that contains 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 2-dimensional Gaussian PDF (9.51), you will realize that the semiaxes of the ellipses are oriented along the eigenvectors of $\boldsymbol{F}_2^{-1}$, that is, they form an angle

$$\tan 2\alpha = \frac{2\rho\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2}, \tag{9.56}$$

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 2-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 $(1.51)^2\pi(\det \boldsymbol{F}_2)^{-1}$. Since an experiment is more constraining when the confidence region is smaller, one can define a simple but useful figure of merit (FOM) as [546]

$$\mathrm{FOM} = \det \boldsymbol{F}_2 \,. \tag{9.57}$$

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. (1.97)]. 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. 9.1.

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 linear 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 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 better.

The only real problem is when there is *almost* a singularity. If the combination is given by $a\theta_1 + b\theta_2 + c\theta_1^2$, then there should be no singularity because of the non-linear term. 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 $\hat{\theta} = a\theta_1 + b\theta_2 + c\theta_1^2$. 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

$$F_{ij}^{(\mathrm{tot})} = F_{ij} + F_{ij}^{(p)} \,. \tag{9.58}$$

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 is 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 $\boldsymbol{F}^{(p)}$) 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)

$$F_{ij}^{(p)} = \frac{\delta_i^m \delta_j^m}{\sigma_m^2} \,. \tag{9.59}$$

So you see that in this case $\boldsymbol{F}^{(p)}$ would be utterly singular but the total $\boldsymbol{F}^{(\mathrm{tot})}$ is not (unless of course $\boldsymbol{F}$ was singular as well for the same parameter, bad luck really).

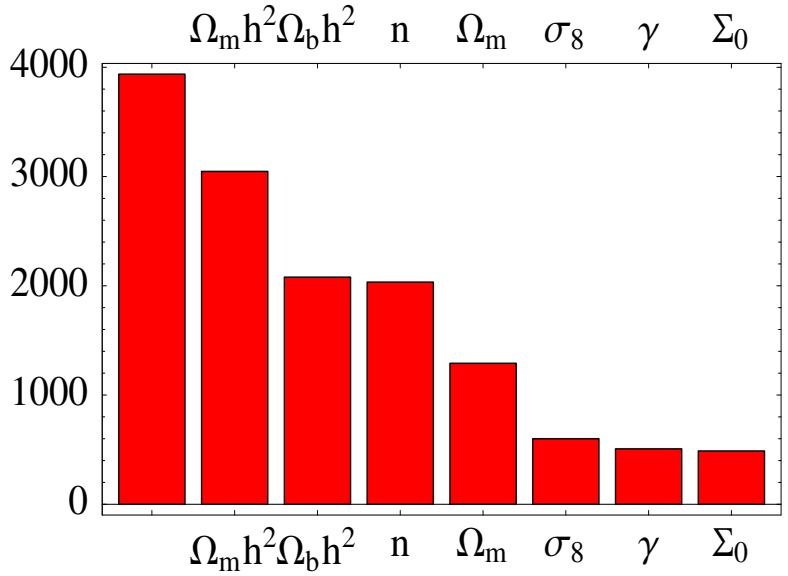

Figure 9.1: The first bar on the right is the FOM for w_0, w_1 with the parametrization (1.97) in a weak lensing experiment assuming all the other cosmological parameters have been fixed to their fiducial values. When additional parameters (listed on the top of the histogram) are marginalized, instead of being held fixed, the FOM reduces. In the figure Ω_m, Ω_b, n, and Σ_0 correspond to $\Omega_m^{(0)}$, $\Omega_b^{(0)}$, n_s, and Σ in our notation, respectively. From Ref. [496].

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 $(\boldsymbol{F}^{-1})_{ii}$ (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 optimist 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 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 circular (or spherical) confidence regions. In some cases this is useful because it reveals hidden properties. 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.

9.4 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 $\delta_{\boldsymbol{k}}$ of a galaxy distribution and their power spectrum calculated for a set of m wavenumbers $\boldsymbol{k}_i$ in some redshift bin $z, z + \Delta z$. Our theory predicts the spectrum $P(k, z; p_i)$ as 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 (see *Lecture notes on Cosmology* for a derivation):

$$\Delta^2_{\boldsymbol{k}} \equiv \langle \delta_{\boldsymbol{k}} \delta^*_{\boldsymbol{k}} \rangle = \langle \delta_{\boldsymbol{k}} \delta_{-\boldsymbol{k}} \rangle = P(\boldsymbol{k}, z) + \frac{1}{n} \,. \tag{9.60}$$

Since the average galaxy density is estimated from the survey itself we have by construction $\langle \delta(x) \rangle = 0$ and therefore $\langle \delta_{\boldsymbol{k}_i} \rangle = 0$ for any $\boldsymbol{k}_i$. The coefficients $\delta_{\boldsymbol{k}_i}$ 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 $\delta_{\boldsymbol{k}_i}$ and the Fisher matrix for the whole $\delta_{\boldsymbol{k}_i}$ 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 $\delta^*_{\boldsymbol{k}} = \delta_{-\boldsymbol{k}}$ 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:

$$L = \frac{1}{(2\pi)^{m/2} \Pi_i \Delta_i} \exp\left[-\frac{1}{2} \sum_i^m \frac{\delta_i^2}{\Delta_i^2} \right] \,, \tag{9.61}$$

(where to simplify notation we write $\Delta_i = \Delta_{\boldsymbol{k}_i}$, $\delta_i = \mathrm{Re}\, \delta_{\boldsymbol{k}_i}$) assuming that the measures at every $\boldsymbol{k}_i$ 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

$$\mathcal{L} = -\ln L = \frac{m}{2} \ln(2\pi) + \sum_i \ln \Delta_i + \sum_i \frac{\delta_i^2}{2\Delta_i^2} \,. \tag{9.62}$$

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. (9.32) the Fisher matrix for a particular z bin is

$$\begin{aligned} F_{\ell m} &= \left\langle \frac{\partial^2 \mathcal{L}}{\partial p_\ell \partial p_m} \right\rangle = \sum \left[\frac{\Delta_{,\ell m}}{\Delta} - \frac{\Delta_{,\ell} \Delta_{,m}}{\Delta^2} - \langle \delta^2 \rangle \left(\frac{\Delta_{,\ell m}}{\Delta^3} - 3 \frac{\Delta_{,\ell} \Delta_m}{\Delta^4} \right) \right] \\ &= \frac{1}{2} \sum_i \frac{\partial \ln P_i}{\partial p_\ell} \frac{\partial \ln P_i}{\partial p_m} \left(\frac{n P_i}{1 + n P_i} \right)^2 \,, \end{aligned} \tag{9.63}$$

[where we used $\langle \delta^2 \rangle = \Delta^2$ from Eq. (9.60)] 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 $k, k + \mathrm{d}k$ and cosine interval $\mathrm{d}\mu$,, i.e. in the Fourier volume $2\pi k^2 \mathrm{d}k \mathrm{d}\mu$. 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 $V_{\rm cell} = (2\pi)^3 / V_{\rm survey}$, so that we have $2\pi k^2 \mathrm{d}k \mathrm{d}\mu / V_{\rm cell} = (2\pi)^{-2} V_{\rm survey} k^2 \mathrm{d}k \mathrm{d}\mu$ modes in the survey volume. The integral form of the Fisher matrix is therefore given by [547, 548]

$$F_{\ell m} = \frac{1}{8\pi^2} \int_{-1}^{+1} \mathrm{d}\mu \int_{k_{\rm min}}^{k_{\rm max}} k^2 \mathrm{d}k \frac{\partial \ln P(k,\mu)}{\partial p_\ell} \frac{\partial \ln P(k,\mu)}{\partial p_m} \left[\frac{n P(k,\mu)}{n P(k,\mu) + 1} \right]^2 V_{\rm survey} \,. \tag{9.64}$$

The factor

$$V_{\rm eff} = \left[\frac{nP(k,\mu)}{nP(k,\mu)+1}\right]^2 V_{\rm survey}\,, \tag{9.65}$$

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 $\boldsymbol{x}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{C}$ distributed according to the likelihood

$$L = \frac{1}{(2\pi)^{n/2}\sqrt{\det \boldsymbol{C}}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})\boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})^t\right]\,, \tag{9.66}$$

where t denotes the transpose. We define the data matrix $\boldsymbol{D} = (\boldsymbol{x}-\boldsymbol{\mu})^t(\boldsymbol{x}-\boldsymbol{\mu})$. Then the covariance matrix is defined in all generality as the expected value of $\boldsymbol{D}$:

$$\langle \boldsymbol{D}\rangle = \boldsymbol{C}\,. \tag{9.67}$$

We can write, up to a constant

$$\mathcal{L} = -\ \ln L = \frac{1}{2}[\ \ln\det \boldsymbol{C} + {\rm Tr}\, \boldsymbol{C}^{-1}\boldsymbol{D}] = \frac{1}{2}{\rm Tr}\,[\ \ln \boldsymbol{C} + \boldsymbol{C}^{-1}\boldsymbol{D}]\,, \tag{9.68}$$

where we used the matrix identity: $\ln\det \boldsymbol{C} = {\rm Tr}\ \ln \boldsymbol{C}$. We suppose now that the theoretical parameters $\boldsymbol{\theta}$ are *both* in $\boldsymbol{\mu}$ and in $\boldsymbol{C}$. The Fisher matrix is then the expected value

$$F_{ij} = \left\langle \frac{\partial^2 \mathcal{L}}{\partial\theta_i \partial\theta_j}\right\rangle \equiv \langle \mathcal{L}_{,ij}\rangle\,, \tag{9.69}$$

To calculate $\langle \mathcal{L}_{,ij}\rangle$ we use the fact that for Gaussian data $\langle \boldsymbol{x}\rangle = \boldsymbol{\mu}$, and consequently

$$\langle \boldsymbol{D}_{,i}\rangle = 0\,, \quad \langle \boldsymbol{D}_{,ij}\rangle = \boldsymbol{\mu}_{,i}\boldsymbol{\mu}^t_{,j} + \boldsymbol{\mu}_{,j}\boldsymbol{\mu}^t_{,i}\,. \tag{9.70}$$

Notice that $\langle \boldsymbol{D}_{,i}\rangle \neq \langle \boldsymbol{D}\rangle_{,i}$. Then we have

$$2\mathcal{L}_{,i} = {\rm Tr}\,[\boldsymbol{C}^{-1}\boldsymbol{C}_{,i}(\boldsymbol{I}-\boldsymbol{C}^{-1}\boldsymbol{D}) + \boldsymbol{C}^{-1}\boldsymbol{D}_{,i}]\,, \tag{9.71}$$

($\boldsymbol{I}$ is the identity matrix) which averages to zero,

$$\langle \mathcal{L}_{,i}\rangle = 0\,. \tag{9.72}$$

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 $\boldsymbol{D}$ since the random variables, the data, are only there, while of course derivatives act only on $\boldsymbol{C}$ and $\boldsymbol{\mu}$ since parameters are only there. To evaluate $\langle \mathcal{L}_{,ij}\rangle$ we notice that all first derivatives $\langle \boldsymbol{D}_{,i}\rangle$ vanish and that $\langle \boldsymbol{I}-\boldsymbol{C}^{-1}\boldsymbol{D}\rangle = 0$. Then we are finally left with [549, 550]

$$F_{ij} \equiv \langle \mathcal{L}_{,ij}\rangle = \frac{1}{2}{\rm Tr}\,\left[\boldsymbol{C}^{-1}\boldsymbol{C}_{,i}\boldsymbol{C}^{-1}\boldsymbol{C}_{,j} + \boldsymbol{C}^{-1}\langle \boldsymbol{D}_{,ij}\rangle\right] = \frac{1}{2}C^{-1}_{\ell m}\frac{\partial C_{mn}}{\partial\theta_i}C^{-1}_{np}\frac{\partial C_{p\ell}}{\partial\theta_j} + C^{-1}_{\ell m}\frac{\partial\mu_\ell}{\partial\theta_i}\frac{\partial\mu_m}{\partial\theta_j}\,, \tag{9.73}$$

(sum over repeated indices) where in the last equality we have written down the full index expression to be more explicit. Equation (9.63) is recovered when $\boldsymbol{\mu} = \boldsymbol{0}$ and $C_{\ell m} = \Delta^2_m \delta_{\ell m}$.

Chapter 10

Cosmology with galaxy clusters

Galaxy clusters occupy a special position in cosmology, since they are the largest gravitationally bounded object in the Universe. They are therefore a bridge between linear scales, where the memory of the initial conditions is still fully traceable, and the strongly non-linear scales in which other non gravitational phenomena take place. Galaxy clusters can be studied by looking at the dynamics of member galaxies, at the X-ray emission of the hot intra cluster gas, at weak and strong lensing, and at the Sunyaev-Zel'dovich effect on the CMB photons. We will discuss some of these topics in this chapter.

10.1 Quick summary

- Galaxy clusters are the largest gravitational bounded object on the Universe
- They can be studied by internal dynamics, strong and weak lensing, X-ray emission of the hot intracluster component, Sunyaev-Zel'dovich effect
- All these probes point to a large amount of dark matter, ten times the baryonic component
- Clusters are a direct probe of cosmology also because their number density depend sensitively on perturbation growth and on volume measurements

10.2 Mass of clusters

Galaxy clusters are groups of hundreds to thousands galaxies within roughly 1 Mpc radius from their center. The closest cluster is Virgo (15 Mpc/h away); the closest rich and regular cluster is Coma (65 Mpc/h). The most regular among them appear as relatively isolated, almost spherical, groups in equilibrium (i.e. without significant subclustering neither in space nor in velocity).

There are 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. We discuss here the first method.

Hydrostatic equilibrium for the ICM gas means that the gradient of the pressure $P_{\rm gas}$ equals the gravitational force:

$$\nabla P_{\rm gas} = -\rho_{\rm gas} \nabla \Phi_N \,, \tag{10.1}$$

where $\rho_{\rm gas}$ is the density and Φ_N is the gravitational potential. Assuming spherical symmetry we obtain

$$\frac{dP_{\rm gas}}{dr} = -\frac{GM\rho_{\rm gas}}{r^2} \,, \tag{10.2}$$

where we have used $\Phi_N = -GM/r$. Assuming the ideal gas equation of state

$$P_{\rm gas} = \frac{N}{V} k_B T = \frac{\rho_{\rm gas}}{\mu m_p} k_B T \,, \tag{10.3}$$

where $\mu \approx 0.6$ is the mean molecular weight for a gas with the expected primordial composition[a] and m_p is the proton mass, we obtain for the mass within a radius r:

$$M(r) = -\frac{r}{G}\frac{k_B T}{\mu m_p}\left(\frac{d\ \ln\rho_{\rm gas}}{d\ \ln r} + \frac{d\ \ln T}{d\ \ln r}\right) . \tag{10.4}$$

This provides a relation between the gas temperature T, the density profile $\rho_{\rm gas}$, 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

$$\rho_{\rm gas} = \frac{\rho_0}{[1+(r/r_c)^2]^{3\beta/2}} , \tag{10.5}$$

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

$$M(r) = \frac{3\beta k_B T(r)}{G\mu m_p}\frac{r^3}{r_c^2+r^2} \approx (1.1\times 10^{14} h^{-1} M_\odot)\beta\frac{T(r) r^3}{r_c^2+r^2} , \tag{10.6}$$

where 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. 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 Vikhlinin et al (2006) is

$$M = M_5\left(\frac{T}{5\,{\rm keV}}\right)^\alpha , \tag{10.7}$$

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 $V_{\rm vir}^2$ scales as $M/R \sim R^2$ i.e. as $M^{2/3}$ and the gas kinetic energy is proportional to the temperature, so that $V_{\rm vir}^2 \propto M^{2/3} \propto T$. Ultimately, a calibration of the mass-temperature relation will be provided by lensing mass estimations. 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 (the Bremmstrahlung spectrum goes like $\exp(-h\nu/k_B T_g)$ and contains often lines from highy ionized iron). From the mass of several clusters one can finally reconstruct the mass function and compare it to the theoretical prediction.

10.3 Baryon fraction[b]

Clusters can contribute to constrain dark energy parameters in another way, first proposed by Sasaki and Pen, 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.

[a] The molecular weight is the mass of a molecule in units of the proton mass, or of 1/12 of ^{12}C. A fully ionized gas can be considered composed of a mixture of "molecules" formed by either nuclei (mostly protons) or electrons. Since the electron mass is negligible, the mean molecular weight is 1/2. Adding a bit of Helium nuclei we obtain $\mu \approx$0.6.

[b] Adapted from Amendola & Tsujikawa, *Dark Energy. Theory and Observations*, CUP 2010.

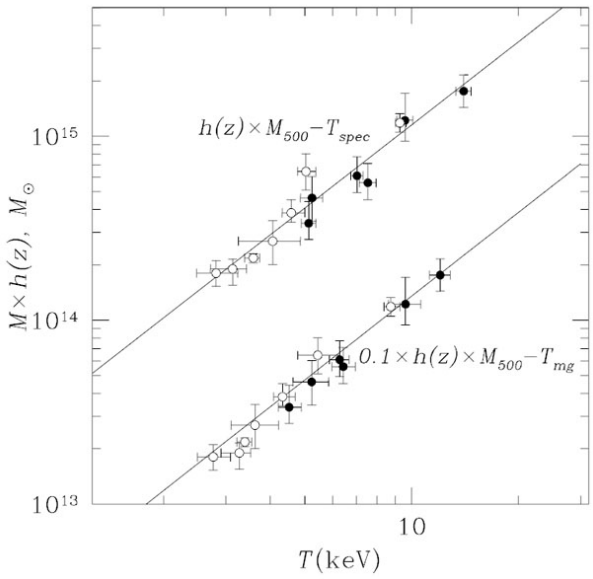

Figure 10.1: X-ray temperature vs. mass in clusters selected by the Chandra satellite (Vikhlinin et al. Astrophys.J.640:691-709, 2006)

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

$$\frac{M_{\rm gas}}{M_{\rm tot}} = \frac{\Omega_b}{\Omega_m} = \text{constant}\,, \tag{10.8}$$

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.

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 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 $M_{\rm gas}$ is in proportion to $\rho_e V$, it follows that $L_X \propto M_{\rm gas}^2/r^3$ or $M_{\rm gas} \propto (L_X r^3)^{1/2}$. We also notice that the X-luminosity is measured by an observed flux $\mathcal{F}_X = L_X/(4\pi d_L^2)$, so we can also write $M_{\rm gas} \propto d_L r^{3/2}$. On the other hand, from the hydrostatic equilibrium condition (10.4), we deduce that the total mass is $M_{\rm tot}(r) \propto r$ (see also Eq. 10.6 at large r), if we assume an isothermal distribution and that $d\,\ln\rho_{\rm gas}/d\,\ln r$ depends weakly on r (which is true for instance for all power-law $\rho_{\rm gas} \sim r^n$). So finally we have

$$\frac{M_{\rm gas}}{M_{\rm tot}} \propto \frac{d_L r^{3/2}}{r} \propto d_L r^{1/2}\,. \tag{10.9}$$

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

$$f_{\rm gas} = \left.\frac{M_{\rm gas}}{M_{\rm tot}}\right|_{<\theta} = A_1 d_L r^{1/2} = A_2 d_L d_A^{1/2} = A_3 d_A^{3/2}\,, \tag{10.10}$$

where A_1, A_2, A_3 are constants or observable quantities like θ and z. Note that we have used the Etherington relation in the last step. The A_i's factors contain a lot of interesting physics but no cosmological parameters, so we are not concerned with them here. Then we see that $f_{\rm gas} d_A^{-3/2}$ is an observable quantity independent of cosmological parameters. So for instance if we have two clusters, taking the ratio of $f_{gas} d_A^{-3/2}$ gives the ratio of their angular diameter distances; so if we know the distance of one cluster we can estimate the distance to

the other. So fitting $f_{\rm gas} d_A^{-3/2}$ 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.

The simple $f_{\rm gas}$ 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. 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 have been derived in Allen et al. (2008) from 42 clusters observed by the Chandra X-ray satellite. The constraint from the $f_{\rm gas}$ test alone gives for the equation of state of dark energy $w_{\rm DE} = -1.14 \pm 0.31$, for flat space and constant $w_{\rm DE}$ (all results here and below are at 1σ). The results in Ettori et al. on a different cluster dataset give $\Omega_m^{(0)} = 0.32^{+0.04}_{-0.05}$ and $w_{\rm DE} = -1.1^{+0.60}_{-0.45}$. In combination with SN Ia and CMB, these constraint tightens to $w_{\rm DE} = -0.98 \pm 0.07$.

10.4 Virial theorem

The mass of clusters can also be estimated by the dynamics of the member galaxies, through the virial theorem and the Jeans equations. Given a system of N particles of equal masses m with position $\mathbf{r}_k$ and momenta $\mathbf{p}_k = m\dot{\mathbf{r}}_k$, we define the quantity

$$G = \sum_k^N \mathbf{r}_k \cdot \mathbf{p}_k \tag{10.11}$$

summing over all particles. We assume the system is at rest, i.e. the average $\langle \dot{\mathbf{r}}_k \rangle$ is zero. The derivative of G is

$$\frac{dG}{dt} = \sum_k (m\dot{\mathbf{r}}_k \cdot \dot{\mathbf{r}}_k + \dot{\mathbf{p}}_k \cdot \mathbf{r}_k) \tag{10.12}$$

$$= \sum_k (m\dot{r}_k^2 + \mathbf{F}_k \cdot \mathbf{r}_k) \tag{10.13}$$

$$= 2T + \sum_k \mathbf{F}_k \cdot \mathbf{r}_k \tag{10.14}$$

where T is the total kinetic energy

$$T = \frac{1}{2} m \sum_k \dot{\mathbf{r}}_k^2 = \frac{1}{2} M \sigma_{rr}^2 \tag{10.15}$$

where $\sigma_{rr}^2 = \frac{1}{N}\sum_k \dot{\mathbf{r}}_k^2 - \langle \dot{\mathbf{r}}_k \rangle^2 = \frac{1}{N}\sum_k \dot{\mathbf{r}}_k^2$ is the particle velocity dispersion and $M = Nm$ the total mass, and $\mathbf{F}_k$ is the force acting on the k-th particle. Now we write this force as the sum of all the forces from the other particles

$$\mathbf{F}_k = \sum_{j=1}^N \mathbf{F}_{jk} \tag{10.16}$$

so that

$$\sum_k \mathbf{F}_k \cdot \mathbf{r}_k = \sum_k (\sum_{j<k} \mathbf{F}_{jk}\mathbf{r}_k + \sum_{j>k} \mathbf{F}_{jk}\mathbf{r}_k) = \sum_k \sum_{j<k} \mathbf{F}_{jk}\mathbf{r}_k + \sum_k \sum_{j>k} \mathbf{F}_{kj}\mathbf{r}_j = \sum_k \sum_{j<k} \mathbf{F}_{jk}\mathbf{r}_k + \sum_j \sum_{k>j} \mathbf{F}_{jk}\mathbf{r}_k = \tag{10.17}$$

$$\sum_k (\sum_{j<k} \mathbf{F}_{jk}\mathbf{r}_k + \sum_{j>k} \mathbf{F}_{jk}\mathbf{r}_k) = \sum_k \sum_{j<k} \mathbf{F}_{jk}(\mathbf{r}_k - \mathbf{r}_j) \tag{10.18}$$

where we have used Newton's third law, $\mathbf{F}_{jk} = -\mathbf{F}_{kj}$ and the identity $\sum_k \sum_{j<k} = \sum_j \sum_{k>j}$. Then if we have a potential that depends only on the distance $r_{jk} = \sqrt{(r_{j,x} - r_{k,x})^2 + (r_{j,y} - r_{k,y})^2 + (r_{j,z} - r_{k,z})^2}$ (eg a

gravitational or Coulomb potential) we have

$$\frac{dr_{jk}}{d\mathbf{r}_{jk}} = \{\frac{dr_{jk}}{d(r_{j,x}-r_{k,x})}, \frac{dr_{jk}}{d(r_{j,y}-r_{k,y})}, \frac{dr_{jk}}{d(r_{j,z}-r_{k,z})}\} \tag{10.19}$$

$$= \frac{\mathbf{r}_k - \mathbf{r}_j}{r_{jk}} \tag{10.20}$$

and therefore

$$\mathbf{F}_{jk} = -\nabla_{r_{jk}} V = -\frac{dV}{dr_{jk}}\frac{dr_{jk}}{d\mathbf{r}_{jk}} = -\frac{dV}{dr_{jk}}\left(\frac{\mathbf{r}_k - \mathbf{r}_j}{r_{jk}}\right) \tag{10.21}$$

Then we have

$$\frac{dG}{dt} = 2T + \sum_k \sum_{j<k} \mathbf{F}_{jk}(\mathbf{r}_k - \mathbf{r}_j) \tag{10.22}$$

$$= 2T - \sum_k \sum_{j<k} \frac{dV}{dr_{jk}}\left(\frac{|\mathbf{r}_k - \mathbf{r}_j|^2}{r_{jk}}\right) \tag{10.23}$$

$$= 2T - \sum_k \sum_{j<k} \frac{dV}{dr_{jk}} r_{jk} \tag{10.24}$$

For a stationary object (often denoted as a *virialized* system), G does not depend on time and then

$$2T = \sum_k \sum_{j<k} \frac{dV}{dr_{jk}} r_{jk} \tag{10.25}$$

For the gravitational potential, $V(r_{jk}) = m^2 r_{jk}^{-1}$ and we obtain

$$2T = m^2 \sum_k \sum_{j<k} r_{jk}^{-1} = -V_{tot} \tag{10.26}$$

where V_{tot} is the total gravitational potential energy. A generalization to $V \sim r^n$ is immediate. So by measuring the total kinetic energy of a system, i.e. in practice its velocity dispersion, we can estimate the total potential energy, which depends essentially on the mass distribution.

For instance, for a homogeneous sphere of uniform density ρ, the potential energy of every shell of thickness dr is $dV = -G(\frac{4\pi}{3}\rho r^3)(4\pi r^2 \rho dr) r^{-1}$ so we have

$$m^2 \sum_k \sum_{j<k} r_{jk}^{-1} = -\int_0^R dV = \frac{16\pi^2 G}{3}\int_0^R \rho^2 r^4 dr \tag{10.27}$$

$$= \frac{16\pi^2 G\rho^2}{3}\frac{R^5}{5} \tag{10.28}$$

$$= \frac{3}{5}\frac{GM^2}{R} \tag{10.29}$$

so that in this case

$$\sigma_{rr}^2 = \frac{3}{5}\frac{GM}{R} \tag{10.30}$$

If we use as velocity dispersion the line-of-sight velocity dispersion (the only observable one) then this value has to be multiplied by 3 since it represents only one of the three components of velocity. Typical values for galaxy clusters are $\sigma_{rr} \approx 1000$ km/sec and $R \approx 1$Mpc so

$$M \approx 10^{45} kg \approx 10^{15} M_{\odot} \tag{10.31}$$

The luminosity of a cluster is around $L = 10^{13} L_{\odot}$ in the visible bands, so the mass-to-light ratio is roughly 100 times the solar one. Even accounting for a large quantity of ionized gas that emits in the X-ray band and not in the visible, the mass of clusters is at least ten times larger than the baryonic mass. This is one of the strongest evidences in favor of the existence of dark matter.

10.5 Sunyaev-Zel'dovich effect

CMB photons passing through a cluster of galaxies have a 1% chance of scattering with a electron of the hot intra-cluster medium, the fully ionized, hot ($10^{7\div8}$K) gas component that is trapped in galaxy clusters by their gravitational field. The scattering is an inverse Compton scattering, meaning that the photons gain, rather than lose, energy. This implies that CMB photons of low energy are boosted to high energy and the black-body spectrum is therefore distorted by roughly 1mK: this is the thermal Sunyaev-Zel'dovich (SZ) effect. Since we still approximate the distorted spectrum by a black-body spectrum, we just express the SZ distortion by a frequency-dependent change in CMB temperature (Fig. 10.2). See e.g. Carlstrom et al., 2002 ARA&A..40..643C for more on cosmology with galaxy clusters.

The relative energy injection to the photons after a single scattering is

$$\frac{k_B T_e}{m_e c^2} \ll 1 \tag{10.32}$$

where T_e is the electron temperature and m_e the electron mass. Taking into account the Thomson cross section σ_T, the total number of scatterings in a electron plasm with number density n_e along the line of sight $d\ell$ is $n_e \sigma_T d\ell$ and the total energy shift is then

$$y = \int n_e \frac{k_B T_e}{m_e c^2} \sigma_T d\ell \tag{10.33}$$

(called Compton y-parameter). Then the change in effective black-body temperature for the distribution of photons as a function of the dimensionless frequency $x = h\nu/k_B T_{CMB}$ is (non-relativistic limit)

$$\frac{\Delta T}{T_{CMB}}|_{SZ} = \left(x \frac{e^x + 1}{e^x - 1} - 4 \right) y \tag{10.34}$$

This function changes sign around 220 GHz. Clusters will then appear on CMB maps as circular shadows below this frequency and as sources above it. Typical values of y are around 10^{-4}. The Planck satellite (2015) has detected more than a thousand SZ clusters in its CMB maps.

An important property of the SZ effect is that it is independent of redshift, so it can be used to map clusters at very high redshifts (of course within the experiment resolution).

The total integrated SZ (ISZ) effect in a cluster is obtained by integrating over the solid angle occupied by the cluster on the sky

$$\Delta_{ISZ} = \int \Delta T d\Omega \approx D_A^{-2} \int n_e \frac{k_B T_e}{m_e c^2} \sigma_T d\ell dA = \frac{k_B \sigma_T}{m_e c^2} D_A^{-2} \int n_e T_e dV \tag{10.35}$$

$$= \alpha N_e T_e D_A^{-2} = \alpha' M T_e D_A^{-2} \tag{10.36}$$

where N_e is the total number of electrons, T_e the density-averaged temperature, α, α' are constants independent of the cluster, M is the total mass of the cluster (assuming is universally proportional to N_e, i.e. to the gas mass) and D_A is the angular diameter distance of the cluster, which enters because by definition $d\Omega = dA/D_A^2$. So if we can estimate N_e or M we can obtain the distance to the cluster. In fact M can be independently estimated through X-ray observations (see Eq. 10.7) so we can use clusters also to map the cosmic expansion independently of supernovae or other distance estimators. The current results are however still not competitive.

An additional SZ effect, called kinetic SZ, is due to the Doppler shift induced on the CMB photons if the cluster is moving with velocity v_{pec} with respect to the CMB; the kinetic SZ is proportional to $v_{pec,\ell}/c$, where $v_{pec,\ell}$ is the projection along the line of sight. The kinetic effect at first order does not distort the black body spectrum and can therefore be distinguished from the thermal one.

Appendix: The Jeans equations

The virial theorem applies to averages over the entire system. If we wish to reconstruct the density profile of galaxies and clusters we need to consider the equilibrium dynamics in more detail.

A flow of particles that are neither destroyed nor created and do not collide against each other is governed by the collisionless Boltzmann equation. Stars in a smooth galactic gravitational potential, or galaxies in galaxy

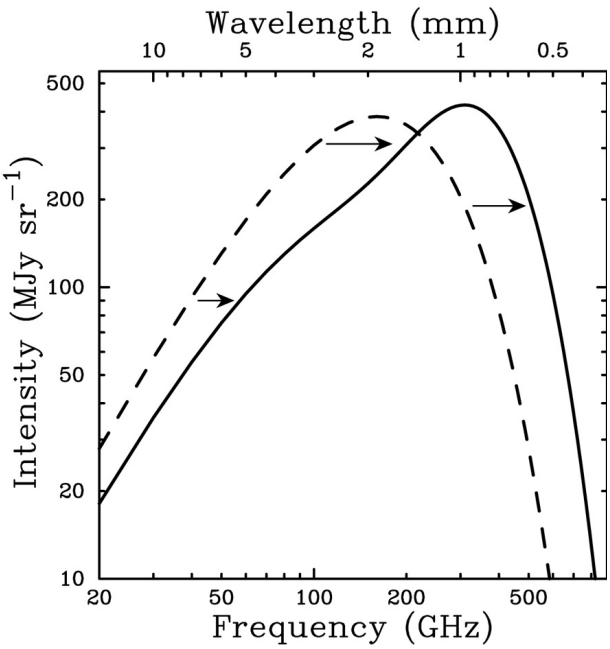

Figure 10.2: The distortion of the Cosmic Microwave Background (CMB) spectrum due to the Sunyaev-Zel'dovich effect (SZ) (solid line), here amplified by a factor of 1000 for readability. The CMB effective temperature decreases at frequencies below approximately 218 GHz and increases at higher frequencies. (From Carlstrom et al., 2002 ARA&A..40..643C)

clusters, satisfy these conditions since they very rarely collide. Let the number dN of particles in a space volume dV and with velocities within a velocity volume $dv_x dv_y dv_z$ be

$$dN = f(t, \mathbf{x}, \mathbf{v}) dx dy dz dv_x dv_y dv_z \tag{10.37}$$

where f is the *distribution function* and the space-velocity volume element is called the phase-space volume. Since the particle number is conserved, the number of particles with velocity v_x entering in a space volume $dxdydz$ in direction x in the time interval dt is $\rho dxdydx = \rho v_x dtdydz$ where ρ is the number density of particles, minus the number exiting from the other side; if the particles move only along x, this has to be equal to the change in the number of particles inside the volume itself, $d\rho dV$. Then we have the continuity equation (the incoming velocity is taken negative)

$$d\rho dV = -d(\rho v_x) dt dy dz \tag{10.38}$$

or

$$\frac{\partial \rho}{\partial t} = -\frac{\partial (\rho v_x)}{\partial x} \tag{10.39}$$

For particle moving in any direction and also with any velocity, we have the general phase-space continuity equation

$$\frac{\partial f}{\partial t} + \sum_{i=1}^{6} \frac{\partial (f \dot{w}_i)}{\partial w_i} = 0 \tag{10.40}$$

where $w = \{x, y, z, v_x, v_y, v_z\} = \{\mathbf{x}, \mathbf{v}\}$ is the phase-space vector of coordinates. Now we have $\dot{v}_i = -\frac{\partial \Phi}{\partial x_i}$ where Φ is the gravitational potential, and

$$\frac{\partial \dot{w}_i}{\partial w_i} = \{\frac{\partial v_i}{\partial x_i}, -\frac{\partial}{\partial v_i}\frac{\partial \Phi}{\partial x_i}\} = 0 \tag{10.41}$$

since the velocity vector is independent of the coordinates and the gravitational potential Φ is independent of the velocities. Then finally

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \sum_{i=1}^{6} \dot{w} \frac{\partial f}{\partial w_i} = 0 \tag{10.42}$$

or the collisionless Boltzmann equation

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla f - \nabla \Phi \cdot \nabla_{\mathbf{v}} f = 0 \tag{10.43}$$

This equation can be written in any system of coordinates, eg cylindrical (useful for disk-like galaxies) or spherical (for spherical galaxies or clusters). In spherical coordinates we have

$$\frac{\partial f}{\partial t} + \dot{r}\frac{\partial f}{\partial r} + \dot{\theta}\frac{\partial f}{\partial \theta} + \dot{\phi}\frac{\partial f}{\partial \phi} + \dot{v}_r \frac{\partial f}{\partial v_r} + \dot{v}_\theta \frac{\partial f}{\partial v_\theta} + \dot{v}_\phi \frac{\partial f}{\partial v_\phi} = 0 \tag{10.44}$$

where we can also write $\dot{r} = v_r$, $\dot{\theta} = v_\theta / r$ and $\dot{\phi} = v_\phi / r \sin\theta$ (here we have defined the velocity components as the projection of the cartesian velocity along the directions r, θ, ϕ, respectively).

The relation of the accelerations in terms of the potential derivatives should be derived by transforming the corresponding cartesian expressions. Here we give directly the equations:

$$\dot{v}_r = \frac{v_\theta^2 + v_\phi^2}{r} - \frac{\partial \Phi}{\partial r} \tag{10.45}$$

$$\dot{v}_\theta = \frac{v_\phi^2 \cot\theta - v_r v_\theta}{r} - \frac{1}{r}\frac{\partial \Phi}{\partial \theta} \tag{10.46}$$

$$\dot{v}_\phi = \frac{-v_\phi v_r - v_\phi v_\theta \cot\theta}{r} - \frac{1}{r\sin\theta}\frac{\partial \Phi}{\partial \phi} \tag{10.47}$$

We assume from now on that the potential is purely spherical. Then we can write

$$\frac{\partial f}{\partial t} + v_r \frac{\partial f}{\partial r} + \frac{v_\theta}{r}\frac{\partial f}{\partial \theta} + \frac{v_\phi}{r\sin\theta}\frac{\partial f}{\partial \phi} + \left(\frac{v_\theta^2 + v_\phi^2}{r} - \frac{\partial \Phi}{\partial r}\right)\frac{\partial f}{\partial v_r} + \left(\frac{v_\phi^2 \cot\theta - v_r v_\theta}{r}\right)\frac{\partial f}{\partial v_\theta} + \left(\frac{-v_\phi v_r - v_\phi v_\theta \cot\theta}{r}\right)\frac{\partial f}{\partial v_\phi} = 0 \tag{10.48}$$

Then we multiply by v_r and obtain

$$v_r \frac{\partial f}{\partial t} + v_r^2 \frac{\partial f}{\partial r} + \frac{v_r v_\theta}{r}\frac{\partial f}{\partial \theta} + \frac{v_r v_\phi}{r\sin\theta}\frac{\partial f}{\partial \phi} + v_r\left(\frac{v_\theta^2 + v_\phi^2}{r} - \frac{\partial \Phi}{\partial r}\right)\frac{\partial f}{\partial v_r}$$
$$+ v_r\left(\frac{v_\phi^2 \cot\theta - v_r v_\theta}{r}\right)\frac{\partial f}{\partial v_\theta} + v_r\left(\frac{-v_\phi v_r - v_\phi v_\theta \cot\theta}{r}\right)\frac{\partial f}{\partial v_\phi} = 0 \tag{10.49}$$

Now we put $\nu(t, \mathbf{x}) = \int f dv_r dv_\theta dv_\phi$, which is then a density independent of velocity, and integrate the above equation over the velocities. We have for instance terms like

$$\int \frac{\partial f}{\partial v_r} dv_r dv_\theta dv_\phi = \frac{\partial}{\partial v_r}\nu(\mathbf{x}) = 0 \tag{10.50}$$

$$\int v_r \frac{\partial f}{\partial r} dv_r dv_\theta dv_\phi = -\int \frac{\partial v_r}{\partial r} f dv_r dv_\theta dv_\phi = -\frac{\partial}{\partial r}\bar{v}_r \tag{10.51}$$

$$\int v_r v_\theta \frac{\partial f}{\partial \theta} dv_r dv_\theta dv_\phi = -\int \frac{\partial (v_r v_\theta)}{\partial \theta} f dv_r dv_\theta dv_\phi = -\frac{\partial}{\partial \theta}\sigma_{r\theta} + \frac{\partial}{\partial \theta}(\bar{v}_r \bar{v}_\theta) \tag{10.52}$$

(we applied integration by parts) where $\bar{v}_r$ is the average of v_r and

$$\sigma_{ij}^2 = \int v_i v_j f dv_r dv_\theta dv_\phi - \bar{v}_i \bar{v}_j \tag{10.53}$$

is the velocity covariance. We now simplify our problem by assuming that: a) the system is stationary (all time derivatives vanish and $\bar{v}_r = 0$; b) the kinematics has spherical symmetry: i.e. there is no rotation ($\bar{v}_\theta = \bar{v}_\phi = 0$) and also $\sigma^2_{r\theta} = \sigma^2_{r\phi} = \sigma^2_{\theta\phi} = 0$; c) finally, we also impose $\sigma^2_{\theta\theta} = \sigma^2_{\phi\phi} \equiv \sigma^2_t$. Then we have the Jeans equation

$$\frac{1}{\nu}\frac{\partial}{\partial r}(\nu\sigma^2_{rr}) + 2\frac{\sigma^2_{rr} - \sigma^2_t}{r} = -\frac{\partial\Phi}{\partial r} = -\frac{GM(r)}{r^2} \tag{10.54}$$

or

$$\frac{1}{\nu}\frac{\partial}{\partial r}(\nu\sigma^2_{rr}) + 2\sigma^2_{rr}\frac{\beta}{r} = -\frac{\partial\Phi}{\partial r} = -\frac{GM(r)}{r^2} \tag{10.55}$$

where

$$\beta = 1 - \frac{\sigma^2_t}{\sigma^2_{rr}} \tag{10.56}$$

is the anisotropy parameter, and it can take values in $-\infty < \beta \leq 1$. Notice that here the mass M includes all sources of gravity while the density ν might refer only to a particular population, e.g. stars or galaxies, and not necessarily to all the matter in the system. If velocities are isotropic, $\sigma^2_{rr} = \sigma^2_t$ we obtain the condition of spherical hydrostatic equilibrium

$$\frac{1}{\nu}\frac{\partial}{\partial r}(\nu\sigma^2_{rr}) = -\frac{GM(r)}{r^2} \tag{10.57}$$

The Jeans equation can be written in a form that directly gives the mass profile as a function of observables

$$M(r) = -\frac{r\sigma^2_{rr}}{G}\left[\frac{d\ \ln\nu}{d\ \ln r} + \frac{d\ \ln\sigma^2_{rr}}{d\ \ln r} + 2\beta(r)\right] \tag{10.58}$$

Notice the similarity to the hydrostatic equation (10.4). However, the function $\beta(r)$ is very difficult to estimate, since we only directly measure the line-of-sight velocity.

Chapter 11

Dark matter

We have mentioned several times the dark matter and have seen some of the evidence in favor of it. This chapter is devoted to the main observational techniques to "observe" dark matter beyond those already mentioned and to a short review of the candidates.

Quick summary

- Dark matter is very likely non-baryonic and composed of free particles.
- Several particle physics models of DM have been proposed but the so-called WIMPs (predicted in supersymmetry) are probably the most widely accepted candidate
- WIMPs of mass 10-100 GeV would have the right amount of density to explain all of the DM
- DM can be detected directly or indirectly, or even produced in accelerators, but so far no conclusive evidence has been reported
- Direct detection might occur when the galaxy halo DM particles cross the Earth and hit a nucleus in the underground detectors
- Indirect detection occurs if DM annihilates when encountering other DM particles in high-density regions (central Milky Way, dwarf galaxies, clusters etc) or decays: we can then observe the products as high energy monochromatic photons, neutrinos, charged leptons.
- Current cold DM models cannot explain easily the galaxy inner profiles and the number of dwarf satellites: this could be a problem related to baryonic physics or require a modification of the standard DM models.

11.1 Dark matter candidates

Dark matter can be composed in principle by any sort of matter that escapes direct observation in every electromagnetic waveband, from γ-rays to radio. However we know from big bang nucleosynthesis and CMB constraints that the baryons can only amount to 5% at most of the cosmic density, while we need around 30% of matter to explain spiral rotation curves, $X-$ray ICM, CMB, cosmic expansion etc. This DM must be cold enough to remain within galaxy and cluster halos: relativistic particles would easily escape the gravitational potentials. Moreover, the DM particles obviously have to have a mass sufficient to provide the required density. So the natural candidates are massive, cold, stable particles that do not interact electromagnetically. Since the existence of an abundant number of micro black holes and in general of unseen compact objects is severely limited by many constraints (for instance, limits coming from microlensing), it is likely that the DM is a gas of such particles freely floating around in galaxies and clusters.

DM particles that feel weak interaction (beside of course gravitational interaction) are called WIMPs (weakly interacting massive particles). In this case, one can estimate their present number density since we know the epoch when weak interacting particles left thermal equilibrium with the other particles (*freezing*) in the early

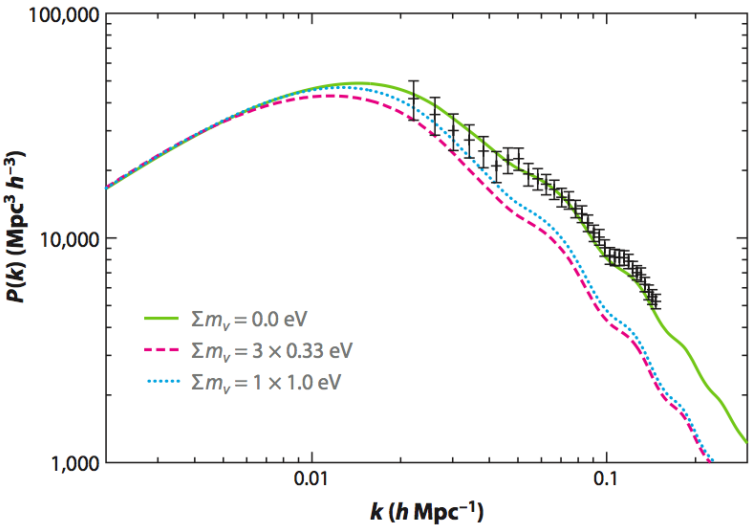

Figure 11.1: Matter power spectrum in models with massless and massive neutrinos (from Wong, Y., Ann.Rev. Nucl.Part.Sci. 61, 2011). The data points are from the 2dF survey (Cole et al. 2005).

universe. It turns out that particles with weak-scale mass around 100 GeV will have automatically more or less the density required to provide all the DM we need: this is called the WIMP miracle. This feature makes the WIMPs the perfect candidate for DM. However, so far no WIMP has ever been captured or produced.

Supersymmetric theories provide a host of new partner particles to the standard model, all with masses above several GeV (otherwise they would have been produced in accelerators; this does not apply to the gravitino however). The lightest supersymmetric particle must be stable because there are no symmetry-conserving particles they can decade to, so supersymmetry provide naturally at least one kind of WIMP, often identified with the neutralino o gravitino or, generically, LSP, "lightest sypersymmetric particle", whichever it be. Most experiments are therefore geared at capturing 10-100 GeV WIMPs passing through the Earth. The fact that the LHC is not finding new physics at these mass scales is generating tension with the WIMP-DM connection scenario.

Another dark matter candidate is the axion, a particle proposed in order to understand why QCD does not violate parity to a higher degree. Its mass is constrained to be very small, of the order of μeV, but it is still of the "cold" type because it has never been in thermal equilibrium with the other particles. Searches for the axion have been unsuccessful so far and large chunks of the interesting parameter space have been excluded.

Many other types of DM have been proposed, also of the "warm" type, for instance sterile neutrinos with mass around keV. One type of DM is certainly realized in nature: neutrinos. Since their mass is constrained to be below roughly 1eV, they are born relativistic (temperature of freezing in the early universe much higher than their mass) and might be becoming non-relativistic just now. Such a small mass implies that they cannot contribute to a large fraction of the DM, at most 1%. Moreover, their large velocity will not allow them to cluster along with baryonic matter so that they will in general decrease the overall matter clustering. In fact, by measuring accurately the matter power spectrum one can infer upper limits to the abundance, and therefore mass, of the neutrinos. In agreement with neutrino oscillation experiments, cosmology constrains the neutrino mass to be at or below the eV level (see Fig. 11.1).

11.2 Direct detection

If dark matter forms a giant halo around the bulge and disk of the Milky Way then the DM particles should cross the Earth and can be in principle detected through their weak interaction with nuclei. The density of the halo should be roughly 100 times the background density, i.e. around 0.3 proton mass per cm^3 so if the DM particles have a mass of say 100 GeV there should be one DM particle every liter of space. If the DM halo is slowly rotating, the Earth will feel a wind of DM particles at a velocity of 220 km/sec (the velocity of the Sun around the Milky Way center), with a little annual modulation along the Earth' orbit around the Sun, for a

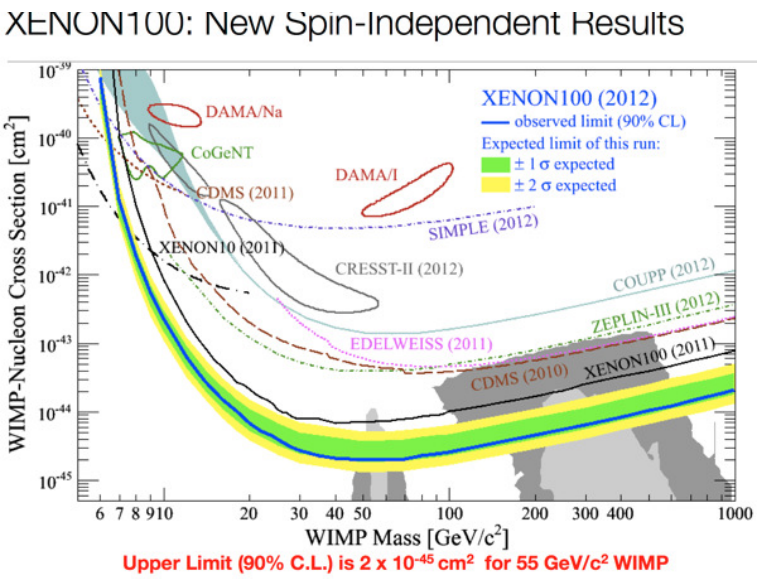

Figure 11.2: Typical exclusion plot cross section versus DM mass collecting the results of many experiments (XENON collaboration, 2012). The closed contours (DAMA, CoGent, CRESST) are positive detections, which appear therefore in contrast with the excluded regions from other experiments. (Credits: http://xenon.astro.columbia.edu/XENON100_Experiment/ and Phys. Rev. Lett. 109, 181301 (2012))

total rate of $10^{4\div5}$ particles per square centimeter per second.

Some of these non-relativistic particles will hit, although with a very low probability, the nuclei of the detectors currently located in various underground laboratories, shielded from the cosmic ray noise by kilometers of rock. A typical event rate could be up to a few events per day per kilogram of detector. Since we expect a good mass range to be around 10-100 GeV, the DM detectors normally employ nuclei with similar mass like Germanium, Silicon, Iodium in order to maximize the recoil. The energy of the recoil is in any case extremely small, up to 100 keV. The recoil can then be measured because, depending on the detector substance, it creates crystal dislocations, ionization, phonons, scintillations.

The main problem is separating true DM events from background, mostly due to natural radioactivity of the rocks and of the detector material itself. Another way is to search for the small annual modulation due to the Earth orbit: the experiment DAMA at Gran Sasso, Italy, claims since several years a positive detection but so far it has not received robust confirmation. Most experiments produced regions of excluded parameter space in the plane cross section versus mass (Fig. 11.2). Low cross sections and low masses are clearly more difficult to constrain and ample regions of parameter space are still unconstrained.

Another form of "almost direct" detection is DM production at high-energy accelerator like LHC at CERN. If the collision of two protons or other particles generates products that do not add up to the original energy, it means some particles have escaped detection. A careful series of tests on backgrounds and on already known elusive particles might in principle lead to the conclusion that the escaped particles are DM candidates.

11.3 Indirect detection

DM particles can be indirectly detected if we can observe signals that are linked to DM decay or self-annihilation. One possibility is that after many interactions some DM particles lose energy and are eventually gravitationally captured by the Sun (or even the Earth). If they accumulate for billion years in the core of the Sun their density might become large enough that their annihilation rate becomes significant. Neutrinos generated by the annihilation will be much more energetic than thermonuclear solar neutrinos and will impact the Earth atmosphere and rock, creating muons. If in a underground laboratory a highly energetic muon traveling *upward* is observed, it cannot come from cosmic rays hitting the atmosphere, so one possibility is that it comes from neutrinos generated in the Earth or Sun's core by DM annihilation. The high-energy neutrinos can also be

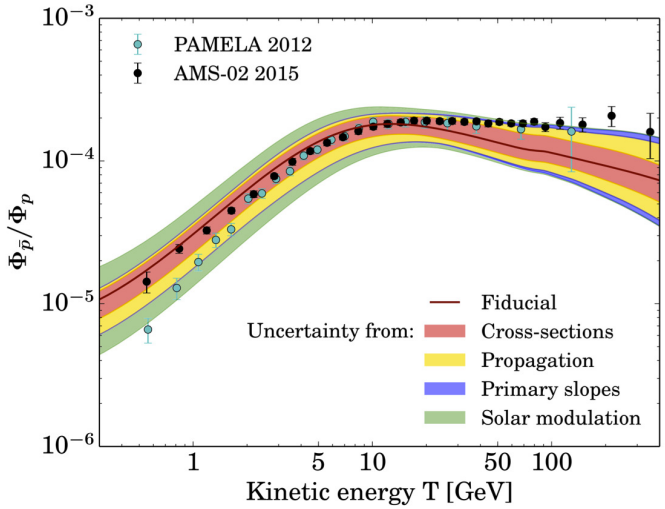

Figure 11.3: Excess of antiproton/proton ratio from the experiments PAMELA and AMS-02. The data are just above the uncertainty band (from Giesen et al. 2015, Journal of Cosmology and Astroparticle Physics, Issue 09, article id. 023, (2015)).

directly captured by neutrino detectors like AMANDA and IceCube.

Another signal from annihilating or decaying DM can be obtained looking for 100 GeV-range γ-ray photons from space, perhaps from regions where we expect a high angular concentration of DM, for instance in the direction of M31 (Andromeda galaxy), the Large Magellanic Cloud, dwarf galaxies in the Local Group, the Milky Way center, galaxy clusters. Since the annihilation is a two-particle process, it is proportional to the DM density squared and this might help revealing the effect and distinguishing it from astrophysical backgrounds (for instance pulsars and supernovae remnants). The photons should have exactly the energy of the DM particles and therefore appear as a sharp line on the continuum coming from high-density regions (redshifted if coming from a moving source). Notwithstanding various temptative detections, so far the satellite Fermi-LAT has excluded WIMP masses up to 100 GeV. Cherenkov detectors, like HESS, MAGIC, VERITAS on Earth can also collect the particle showers generated from the impact of the high-energy photons with the atmosphere.

DM annihilation/decay products could also be pairs of charged fermions (protons, electrons and their antiparticles). The antiparticles will generate an excess over the standard astrophysical background particle/antiparticle ratio and could be revealed by the satellite PAMELA and the instrument AMS on board the ISS. Here again several detection claims have arisen but none has proven to be conclusive evidence of new physics (see Fig. 11.3).

11.4 The problems of the cold dark matter

Although the CDM is widely regarded as the most likely form of dark matter, it is not without problems. The main problem is the excess of small substructure we see (and expect to see) in N-body simulations with respect to what we observe in galaxies (Fig. 11.4). The high-mass end disagreement is not so critical because large halos will in general host groups or clusters instead of individual galaxies. Since CDM does not dissipate efficiently because weakly interacting, and because the theoretical power spectrum has no natural cut-off scale, the CDM lumps formed under gravitational instability live practically forever as satellites within the halos of large galaxies as ours. But around the Milky Way we observe only a few dozen dwarf galaxies, although we cannot easily estimate how many (perhaps hundreds) we are missing due to low surface brightness. Of course we do not know whether every DM lump should contain stars or instead just remain dark either because not enough baryons are collected or because they remain too hot to collapse into stars. So in fact the DM subclustering problems is probably just a manifestation of our uncertainties about baryonic physics, that is, of exactly how and when

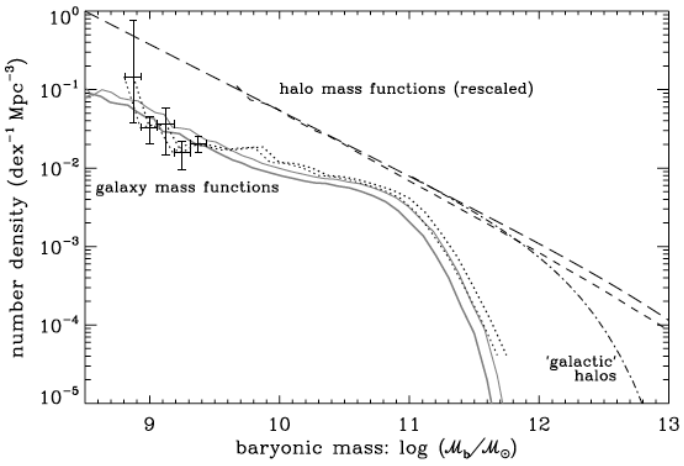

Figure 11.4: Halo mass function from N-body simulations (dahed lines) compared with the galaxy mass/luminosity function from real data (continuous and dotted lines). The discrepancy means that there is a non-linear relation between halos and galaxies. In particular, at the small mass end, the simulations show much more halos than galaxies (from Baldry et al., *On the galaxy stellar mass function, the mass–metallicity relation and the implied baryonic mass function*, Mon. Not. R. Astron. Soc. 388, 945–959 (2008)).

exactly stars finally form.

The clustering problem of DM arises at small scales and its consequence is not just an excess of satellites but also a inner halo profile that seem much steeper (cuspy) that the one that best fits galaxy rotation curves: this is called the cusp-core problem. Here again the DM evidence comes from N-body simulations, well fitted by the steep NFW profile, while observations refer to luminous sources, so baryonic physics might very likely play an important role. Alternatively, DM could depart from the simplest theoretical expectations and allow for a stronger self-interaction that introduces some form of dissipative friction that erases substructure.

Chapter 12

The Cosmic 21cm Background

Since observations of the Cosmic Microwave Background were able to accurately constrain cosmology, further cosmological backgrounds come into focus. This chapter introduces the most prominent 'new' background, the 21cm cosmic background of neutral hydrogen.

Quick summary

- Through the forbidden spin-flip transition of neutral hydrogen radiation is emitted at 21cm rest-frame wavelength.
- The 21cm signal of neutral hydrogen is measured as a brightness temperature relative to the CMB background; depending on the properties of the gaseous medium and other radiative backgrounds the signal is seen in absorption or emission.
- The 21cm brightness temperature depends on the so-called spin temperature, density and ionisation state as well as cosmology.
- During reionization the before neutral medium becomes ionised (again) through the radiation of the first galaxies.
- The 21cm signal is a tracer of the large-scale structure. During reionization it traces the neutral hydrogen distributed in the intergalactic medium (IGM), after reionisation hydrogen shielded in galaxies.

12.1 Fundamentals of the 21cm signal of neutral hydrogen

Hydrogen is the most abundant element in the Universe. After the end of recombination around $z = 1000$, the cosmic hydrogen remained neutral until the UV photons emitted by the first stars and galaxies began to ionize it at redshifts around 10. Investigating this Epoch of Reionization (EoR) is extremely useful to understand the early evolution of stars and galaxies. Furthermore, measuring the growth of structures during reionization is probably one of the most important goals for today's cosmology in order to close the gap between the CMB and lower-redshift studies such as galaxy clustering. The EoR ends around $z \approx 6$ but pockets of neutral hydrogen remain inside or nearby structures, e.g. along filaments, protected by ionizing radiation by their own high density (called self-shielding). Since these clouds are associated with large scale structure, they can constitute a valuable way to reconstruct matter fluctuations up to redshifts of 5-6. Fig. 12.1 presents a sketch that positions the process of reionization in our cosmic timeline.

Neutral cold hydrogen has no energy levels in the optical nor in the radio, the two best windows for ground observations, see Fig. 12.2. However, the two possible electron spin states, parallel or antiparallel with respect to the proton, have a slightly different energy, the parallel state being higher by roughly $E_{01} = h\nu = 5 \cdot 10^{-6}$ eV, as first pointed out by Van de Hulst in 1942. A transition is accompanied by absorption or emission of a photon of wavelength 21 cm, or ≈ 1.42 GHz frequency. At redshift z, this hyperfine line, a forbidden spin-flip transition, becomes $21(1 + z)$ cm, of the order of meters. Being a sharp monochromatic radio wavelength, the

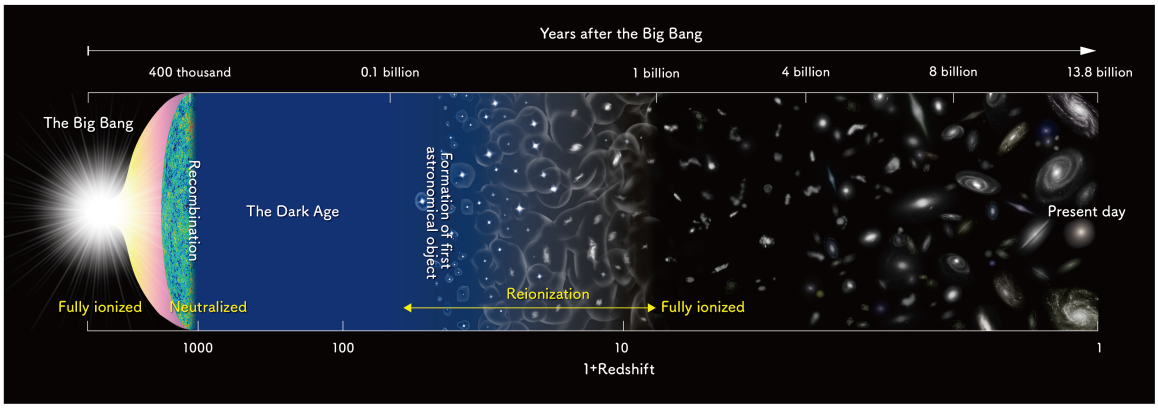

Figure 12.1: A timeline showing the evolution of the Universe. As the first stars and galaxies began to form, they illuminated the dark ages and reionized the Universe. (Credit: NAOJ).

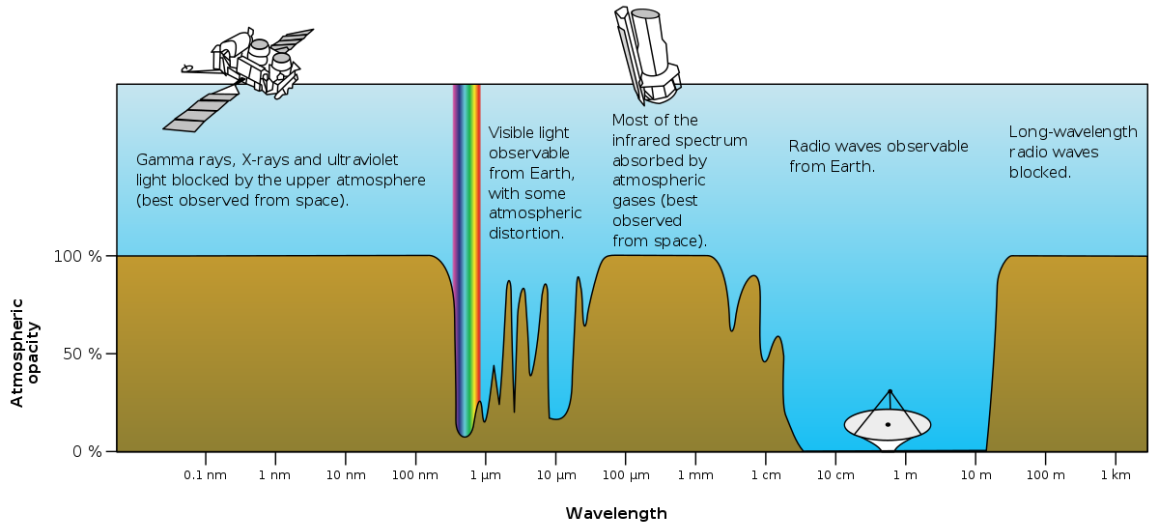

Figure 12.2: Windows in the atmospheric absorption. (Image: NASA)

21 cm line can easily be seen through dust and allows both very accurate galactic rotation curves and precise redshift measurement at very high redshifts. The presentation below can only roughly sketch the cosmological evolution of the 21cm signal of neutral hydrogen, due to the complexity of the problem. Much more information can be found e.g. in Furlanetto et al., astro-ph/0608032.

In the following we use subscript 0 for the anti-parallel singlet state (multiplicity $g_0 = 1$) and subscript 1 for the parallel triplet state ($g_1 = 3$). Neutral hydrogen before and during the EoR is excited mainly by the CMB radiation with temperature $T_\gamma = T_0(1+z)$, where $T_0 = 2.7\,\mathrm{K}$, which is hot enough to saturate the higher levels, $T_\gamma \gg \left(\frac{k_\mathrm{B}}{E_{01}}\right)$, where we define a spin temperature T_S as the temperature one should have in thermodynamic equilibrium to maintain a density of population n_0, n_1 in the two states. We have then

$$\frac{n_1}{n_0} = \frac{g_1}{g_0} e^{-\frac{E_{01}}{k_\mathrm{B} T_\mathrm{S}}} \approx \frac{g_1}{g_0} = 3 \tag{12.1}$$

because the spin temperature turns out to be close to T_γ. That is, three atoms out of four will be in the excited state. We can also write the total number density of HI atoms as $n_\mathrm{HI} = n_0 + n_1 = 4n_0$.

The 21cm signal at each epoch in emission or absorption is characteristic by the state of the gaseous medium, radiative backgrounds such as the CMB, and sources of ionizing radiation such as stars and galaxies. To estimate

T_S three processes are important: 1) absorption and stimulated emission of CMB photons; 2) collisions with other hydrogen atoms, free electrons, and protons; and 3) Lyman-α photons (UV) scattering. The coupling of the excitation or spin temperature of neutral hydrogen to the Lyman-α background is also known as Wouthuysen-Field coupling. We denote with B_{01} the coefficient for absorption (from state 0 to state 1) and with B_{10} the emission one ($1 \to 0$), and with C_{10}, P_{10} the de-excitation rates per atom from collisions and UV scattering, and with C_{01}, P_{01} the corresponding excitation rates. At equilibrium we have then

$$n_1(C_{10} + P_{10} + A_{10} + B_{10}I_\gamma) = n_0(C_{01} + P_{01} + A_{01} + B_{01}I_\gamma). \tag{12.2}$$

The Einstein coefficients depend on the quantum mechanical properties of the atom. One can use the standard relation among Einstein coefficients (see e.g. K. Lang, Astrophysical Formulae, Eq. 2-27)

$$B_{01} = \frac{3c^3 A_{10}}{8\pi h\nu^3}, \tag{12.3}$$

where the Einstein coefficient are furthermore related by $B_{01}/B_{10} = g_1/g_0$. Finally A_{10} is the spontaneous emission coefficient, estimated to be $A_{10} = 2.85 \times 10^{-15}\mathrm{s}^{-1}$ (i.e., one spin flip every 10 million years!). Now we can replace I_γ with T_γ using the Rayleigh-Jeans relation. Then we can invoke equilibrium and write

$$\frac{C_{01}}{C_{10}} = \frac{g_1}{g_0} e^{-\frac{h\nu}{k_\mathrm{B}T_\mathrm{K}}} \approx 3\left(1 - \frac{h\nu}{k_\mathrm{B}T_\mathrm{K}}\right), \tag{12.4}$$

where T_K is the kinetic temperature, and similarly define an effective "color temperature" of the UV ionizing field T_c via

$$\frac{P_{01}}{P_{10}} \equiv 3\left(1 - \frac{h\nu}{k_\mathrm{B}T_\mathrm{c}}\right). \tag{12.5}$$

Putting all this into (12.2) we find

$$T_\mathrm{S}^{-1} = \frac{T_\gamma^{-1} + x_\mathrm{c}T_\mathrm{K}^{-1} + x_\alpha T_\mathrm{c}^{-1}}{1 + x_\mathrm{c} + x_\alpha} \tag{12.6}$$

where

$$x_\mathrm{c} = \frac{C_{10}}{A_{10}} \frac{h\nu}{k_\mathrm{B}T_\gamma} \tag{12.7}$$

$$x_\alpha = \frac{P_{10}}{A_{10}} \frac{h\nu}{k_\mathrm{B}T_\gamma} \tag{12.8}$$

are the coupling coefficients for collisions and scattering, respectively. In the limit in which the couplings are small, this becomes $T_\mathrm{S} = T_\gamma$ as one should expect. In most cases we can assume $T_c \to T_\mathrm{K}$. The exact estimation of T_S can be achieved only numerically. Depending on which process dominates, one can have T_S larger or smaller than T_γ.

12.2 Evolution during the Epoch of Reionization

To observe the 21cm signal during the EoR, one compares at a given redshift z the radiation intensity of the 21cm line with the CMB temperature. Depending on whether the spin temperature T_S is higher or smaller than T_γ, one sees absorption or emission. In this way, it is possible to create maps of HI emission or absorption at several z's. Current radio interferometers statistically measure the power spectrum of the 21cm signal during reionization and are working towards measuring the cosmological signal among foregrounds. The task of imaging the EoR is still to be achieved, but the ambitious large-scale project of the Square Kilometre Array (SKA) is currently set up in South Africa and Australia.

Qualitatively, different regimes for the evolution of the redshifted 21 cm signal are depicted in Fig. 12.3. We briefly describe the evolution here. Starting with the Dark Ages after recombination and before the first galaxies form, the gas temperature T_K is coupled to the CMB temperature T_γ, and therefore the spin temperature T_S

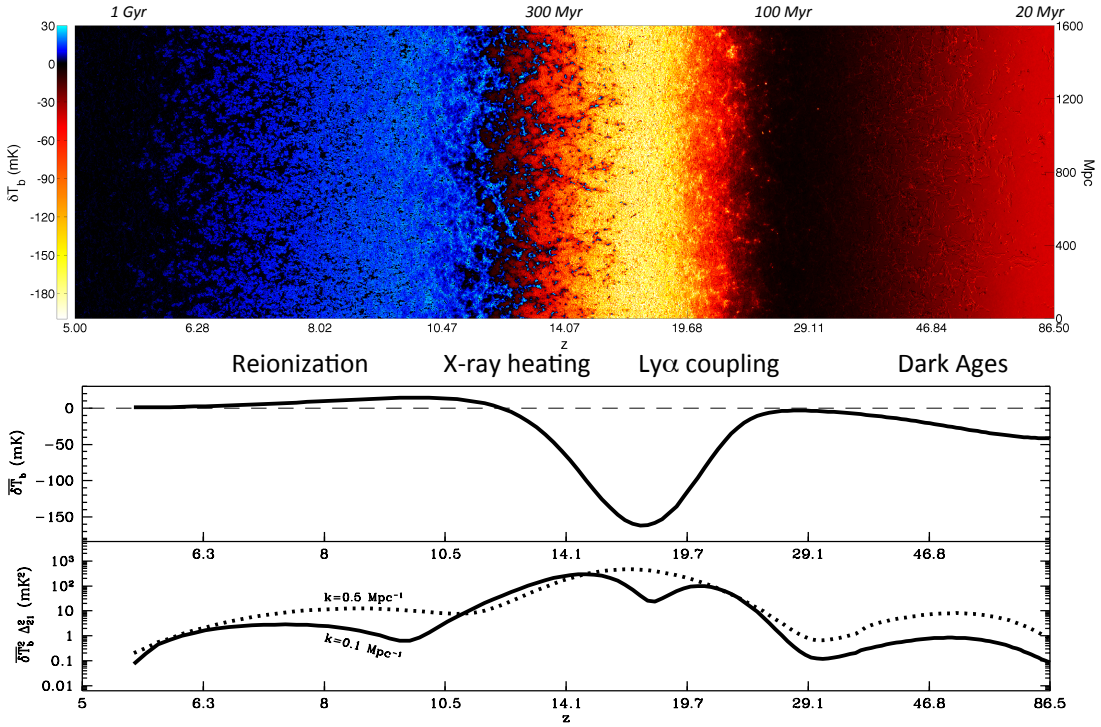

Figure 12.3: 21cm signal from neutral hydrogen (from Mesinger, Greig, Sobacchi, 2016, MNRAS 459, 3, image publicly available here: http://homepage.sns.it/mesinger/EOS.html).

is effectively coupled collisionally to T_γ in a sufficiently dense and neutral medium, setting $T_S = T_\gamma$. Around redshifts of $30 \lesssim z \lesssim 200$ the gas cools adiabatically and begins to thermally decouple from the CMB with $T_S < T_\gamma$, resulting in a shallow absorption feature. When the density becomes low enough for collisional coupling between T_S and T_γ to be negligible, so that no more absorption takes place, then radiative coupling sets $T_S = T_\gamma$ again. There is no 21cm signal, neither in absorption nor in emission. When the first galaxies form at $z \gtrsim 30$, they start to emit both Lyα (UV) and X-ray radiation. At lower emissivities Lyα coupling occurs first and the temperature of the cold gas and the spin temperature get coupled, so that $T_S < T_\gamma$, resulting in a deep absorption feature. Fluctuations in emissivity and density are the most important now, until the Lyα coupling saturates. An increasing X-ray temperature heats the intergalactic medium (IGM) above the CMB temperature, so that neutral hydrogen becomes visible in 21cm emission at redshift between 10-15. Fluctuations in the 21cm line are sourced by temperature fluctuations, and when the gas is heated everywhere, increasingly by density and ionization fluctuations. When the gas temperature reaches the post-heating regime of $T_K \gg T_\gamma$, with $T_S \sim T_K$, the dependence of the 21cm brightness temperature on the spin temperature becomes negligible. As the ionizing UV and X-ray radiation proceeds with reionizing the initial neutral medium, the 21cm signal slowly decreases, until the Universe is (almost) fully ionized around a redshift of ~ 6.

Let us now imagine a background radiation field T_γ whose brigthness or specific intensity is I_γ and a gas of HI at spin temperature T_S and brigthness I_S. (Brightness is related to the flux density S_ν as $S_\nu = \int I_\nu \cos\theta d\Omega$, while the total flux over all frequencies is $S = \int S_\nu d\nu$.) Since the 21cm wavelength is much longer than the peak CMB radiation, the Rayleigh-Jeans region of the black body spectrum is a good approximation, and the brightness and the temperature are simply proportional, $T = Ic^2/2k_B\nu^2$. We can write the observed temperature at frequency ν by solving the radiative transfer equation along a radial coordinate s,

$$\frac{dI_\nu}{ds} = -\alpha(\nu)I_\nu + \alpha(\nu)I_S \tag{12.9}$$

which simply says that the intensity of a radiation field (here identified with the CMB radiation) along propagation decreases because of absorption and increases because of the medium emissivity I_S, assumed not to depend

on s. We define the optical depth

$$\tau = \int \alpha(\nu) \mathrm{d}s \tag{12.10}$$

where $\alpha(\nu)$ is the absorption coefficient at frequency ν and $\mathrm{d}s$ integrates along the line of sight. Note that all frequencies in what follows are rest frame values, not redshifted ones. The solution for the observable brightness I_B is

$$I_\mathrm{B} = I_\mathrm{S}\left(1 - e^{-\tau}\right) + I_\gamma e^{-\tau}. \tag{12.11}$$

Since in the Rayleigh-Jeans regime the temperature T is proportional to brightness I, we can also write

$$T_\mathrm{B} = T_\mathrm{S}\left(1 - e^{-\tau}\right) + T_\gamma e^{-\tau} \tag{12.12}$$

where also T_B scales with redshift as $(1+z)$. Intuitively, if the material is optically thick, $\tau \gg 1$, we see the spin temperature, if optically thin, $\tau \ll 1$, we see the CMB background. This relation can also be written as

$$T_\mathrm{B0}(1+z) = T_\mathrm{S}\left(1 - e^{-\tau}\right) + T_0(1+z)e^{-\tau} \tag{12.13}$$

and therefore

$$\delta T_\mathrm{B} \equiv T_\mathrm{B0} - T_0 = \frac{T_\mathrm{S} - T_\gamma}{1+z}\left(1 - e^{-\tau}\right). \tag{12.14}$$

The quantity δT_B, called brigthness offset temperature, is the main observable: the difference between the observed and the CMB temperature at the desired frequency. Now we need to estimate τ, while T_S follows Eq. 12.6.

The absorption coefficient gives the loss of energy per unit frequency of a radiation field resulting from absorption of photons $(0 \to 1)$ minus stimulated emission $(1 \to 0)$. Since τ is dimensionless, $\alpha(\nu)$ has dimensions m^{-1}. We need to introduce then the Einstein coefficients B_ab for the transition a $\to$ b: the Einstein coefficient is the probability per unit time per unit spectral energy density, where the spectral energy is the energy density per unit frequency, so B_ab has units $J^{-1}m^3s^{-2}$. The rate (i.e. the probability per unit time) of absorption/emission is then given by $B_\mathrm{ab}h\nu n\phi(\nu)$, where n is the number density of absorbers (atoms), $\phi(\nu)$ the line profile normalized so that $\int \phi(nu)d\nu = 1$ and $h\nu n\phi(\nu)d\nu$ is the radiation energy density in the range $d\nu$. Dividing the rate by c, one gets $\alpha_\mathrm{ab} = B_\mathrm{ab}h\nu n\phi(\nu)/c$, the probability of transition per unit distance. This quantity, summing all possible absorptions and substracting all possible emissions, defines the net absorption coefficient: once integrated over distance, it gives the optical depth.

Denoting again with B_{01} the coefficient for absorption (from state 0 to state 1) and with B_{10} the emission one $(1 \to 0)$, the absorption coefficient is then given by the expression

$$\alpha(\nu) = \frac{h\nu}{c}\left(B_{01}n_0 - n_1 B_{10}\right)\phi(\nu). \tag{12.15}$$

The Einstein coefficients are related by $B_{01}/B_{10} = g_1/g_0$ and thus

$$\alpha(\nu) = \frac{h\nu}{c}n_0 B_{01}\left(1 - \frac{n_1}{n_0}\frac{B_{10}}{B_{01}}\right)\phi(\nu) = \frac{h\nu}{c}n_0 B_{01}\left(1 - e^{-\frac{h\nu}{k_B T_S}}\right)\phi(\nu) \approx \frac{h\nu}{c}n_0 B_{01}\frac{h\nu}{k_B T_S}\phi(\nu) \tag{12.16}$$

(we replaced E_{01} with $h\nu$). Therefore we find the dimensionless optical depth

$$\tau = \int \alpha(\nu)\phi(\nu)ds = \frac{h\nu}{c}B_{01}\frac{h\nu}{k_B T_S}\phi(\nu)\int n_0 ds = \frac{h\nu}{c}B_{01}\frac{h\nu}{k_B T_S}\frac{N_\mathrm{HI}}{4}\phi(\nu) \tag{12.17}$$

where $N_\mathrm{HI} = \int n_\mathrm{HI}\mathrm{d}s$ is the *column density* of HI atoms.

The line profile is broadened, in general, by thermal motion, Doppler shift and cosmological recession. Here the dominating effect is the latter one, in which a frequency ν is spread over a range $\frac{\Delta\nu}{\nu} \approx \frac{\Delta V}{c}$ where ΔV is the velocity spread given by $\Delta V = sH(z)$. Therefore we can approximate $\phi(\nu) \approx 1/\Delta\nu \approx c/(sH(z)\nu)$. Then we have

$$\tau = A\frac{x_\mathrm{HI}n_\mathrm{H}}{T_\mathrm{S}H(z)} \tag{12.18}$$

where $A = \frac{3c^3 h A_{10}}{32\pi\nu^2 k_{\rm B}}$ and we write the column density across a distance s as $N_{\rm HI} = x_{\rm HI} n_{\rm H} s$, $x_{\rm HI}(z)$ being the neutral hydrogen fraction and $n_{\rm H}(z)$ the density of hydrogen atoms. We can use now the relation

$$n_{\rm H} = (1+\delta_{\rm b})\,\rho_{\rm b} = (1+\delta_{\rm b})\,\frac{3H_0^2}{8\pi G m_{\rm p}}\Omega_{\rm b,0}\,(1-Y_{\rm He})\,(1+z)^3 \tag{12.19}$$

(the factor of $m_{\rm p}$, mass of the proton, is needed to convert the energy density $\rho_{\rm b}$ into the number density $n_{\rm H}$) where $\delta_{\rm b}$ is the fluctuation in the baryon content and $Y_{\rm He} \approx 0.245$ is the fraction in mass in Helium (all the rest being hydrogen). We use also $H(z) = H_0 \Omega_{\rm m,0}^{1/2}(1+z)^{3/2}$, and finally we obtain from (12.14) and $\tau \ll 1$ the brigthness

$$\delta T_{\rm B} \approx \frac{3}{8\pi G m_{\rm p}} A \left(1 - \frac{T_\gamma}{T_{\rm S}}\right) \frac{x_{\rm HI}(1+\delta_{\rm b})\Omega_{\rm b,0}\,(1-Y_{\rm He})\,H_0^2(1+z)^2}{T_{\rm S} H(z)} \tag{12.20}$$

$$= \frac{3H_0 A\,(1-Y_{\rm He})}{8\pi G m_{\rm p}} \left(1 - \frac{T_\gamma}{T_{\rm S}}\right) x_{\rm HI}(1+z)^{1/2} \frac{\Omega_{\rm b,0}}{\Omega_{\rm m,0}^{1/2}} \tag{12.21}$$

$$\approx 27 x_{\rm HI}\,(1+\delta_{\rm b}) \left(1 - \frac{T_\gamma}{T_{\rm S}}\right) \left(\frac{1+z}{10}\frac{0.15}{\Omega_{\rm m,0}h^2}\right)^{1/2} \left(\frac{\Omega_{\rm b,0}h^2}{0.023}\right) mK \tag{12.22}$$

where the last expression allows an immediate estimation of the order of magnitude, since all factors are of order unity. This expression can be further corrected by multiplying it by $H/({\rm d}v/{\rm d}r + H)$, where ${\rm d}v/dr$ gives the contribution to the line width due to the peculiar velocities along the line of sight. If $T_{\rm S} > T_\gamma$, the brigthness offset $\delta T_{\rm B}$ is positive (emission). Detection of $\delta T_{\rm B}$ can therefore map the history of reionization via ionization fraction $x_{\rm HII}(z) = 1 - x_{\rm HI}(z)$, measure a combination of $\Omega_{\rm b}$ and $\Omega_{\rm m}$, and map the spatial and time structure of the fluctuations in the baryon component $\delta_{\rm b}$. When measuring fluctuations, the 21cm fluctuations $\delta_{21}(\mathbf{x}, \mathbf{z})$ for the 21 cm brightness temperature offset at position $\mathbf{x}$ and redshift z are calculated as

$$\delta_{21}(\mathbf{x}, \mathbf{z}) = \frac{\delta T_{\rm B}(\mathbf{x}, \mathbf{z})}{\bar{T}_{21}(z)} - 1\,, \tag{12.23}$$

with average 21 cm temperature $\bar{T}_{21} = \delta\bar{T}_{\rm B}(z)$; analogous for fluctuations in surface brightness. The dimensionless 21 cm power spectrum is defined as

$$\tilde{\Delta}_{21}(k) = \frac{k^3}{(2\pi^2 V)} \left\langle |\delta_{21}|^2 \right\rangle_k\,, \tag{12.24}$$

and the dimensional power spectrum can be expressed as $\Delta_{21}(k) = \bar{T}_{21}^2 \tilde{\Delta}_{21}(k)$. The power spectrum is sensitive to the reionization and cosmological model parameters, especially as it measures model-dependent behaviour over a large range of scales. The bottom panel of Fig. 12.3 shows the example of a power spectrum prediction at fixed scale k as a function of redshift.

12.3 The 21cm-galaxy cross-correlation during reionization

Bringing the 21cm signal together with galaxy surveys helps to (1) remove foregrounds and systematics, (2) to break degeneracies between model parameters, and (3) constrain the properties of the galaxies that are too faint to be observed with galaxy surveys. We will therefore investigate the power of cross-correlations between the 21cm signal and galaxies in more detail now.

The Lyα line (transition from $n = 2$ to $n = 1$ orbitals of hydrogen) is a prominent line in so-called Lyman-alpha emitters (LAE) that allows for precise redshift estimates. The Lyman-α rest frame wavelength of 1215 angstroms ($\approx$100 nm) falls in the UV region. When redshifted by a factor of $1 + z$ equal to 3 to 8, therefore, the line falls in the visible to near infrared part of the electromagnetic spectrum and can be detected from the ground. For reionization scenarios and LAE models, the 21cm-LAE cross correlations exhibit signatures sensitive to the neutral fraction $\langle\chi_{\rm HI}\rangle$. This largely stems from the large-scale anti-correlation between the neutral 21cm signal and the LAEs. As LAEs are located in ionised regions the corresponding cross-correlations also traces the size of ionised regions around LAEs. Here we will showcase fundamental analytical relations for the 21cm-LAE cross-correlation signal.

We use the following definition for the 21cm-LAE cross correlation function $\xi_{21,\mathrm{LAE}}$

$$\xi_{21,\mathrm{LAE}}(\mathbf{r}) = \langle \delta_{\mathbf{21}}(\mathbf{x}+\mathbf{r})\ \delta_{\mathrm{LAE}}(\mathbf{x})\rangle \tag{12.25}$$

which gives us the excess probability as compared to random distributions of both 21cm signal and LAE at position $\mathbf{x}$ and for distance between both of $\mathbf{r}$. Here the galaxy, or LAE, overdensity is taken smoothed as

$$\delta_{\mathrm{LAE}}(\mathbf{x}) = \frac{N_{\mathrm{LAE}}(\mathbf{x})}{\bar{N}_{\mathrm{LAE}}} - 1, \tag{12.26}$$

with number of LAEs per voxel $N_{\mathrm{LAE}}(\mathbf{x})$ and mean number $\bar{N}_{\mathrm{LAE}}$, similar to the 21cm fluctuations defined in Eq. 12.23.

The cross correlation amplitude at LAE positions
We start by calculating the cross correlation amplitude for distance $r = 0$ and, therefore, at the LAE positions themselves. It reads

$$\begin{aligned}\xi_{21,\mathrm{LAE}}(r=0) &= \frac{1}{N_{\mathrm{bins}}}\sum_{\mathbf{x}}\delta_{21}(\mathbf{x})\ \delta_{\mathrm{LAE}}(\mathbf{x}) &(12.27)\\ &\simeq \frac{1}{N_{\mathrm{bins}}}\sum_{\mathbf{x}}\left(1-\frac{T_\gamma}{T_{\mathrm{S}}(\mathbf{x})}\right)x_{\mathrm{HI}}(\mathbf{x})\frac{\rho_{\mathrm{g}}(\mathbf{x})}{\langle\rho_{\mathrm{g}}\rangle}\delta_{\mathrm{LAE}}(\mathbf{x}) &(12.28)\end{aligned}$$

where $\rho_{\mathrm{g}}(\mathbf{x})$ is the local gas density at position $\mathbf{x}$, and $\langle\rho_{\mathrm{g}}\rangle$ is the mean gas density across the measured region. On the one hand, we see from equation 12.26 that δ_{LAE} adopts only positive values at LAE locations and remains otherwise negative with a value of -1. On the other hand the 21cm signal vanishes in ionized regions ($x_{\mathrm{HI}} = 0$). Since LAEs are located in ionized regions, the 21cm signal is approximately fully absorbed at LAE locations, $\delta_{21} = 0$. Hence, the only regions contributing to the 21cm-LAE cross correlations $\xi_{21,\mathrm{LAE}}(\mathbf{r}=0)$ are the neutral regions where no LAEs can be found. Hence, the 21cm-LAE cross correlations at very small scales can be re-calculated as

$$\begin{aligned}\xi_{21,\mathrm{LAE}}(r=0) &= \frac{1}{N_{\mathrm{bins}}}\left[\sum_{\mathbf{x}_{\mathrm{HI}}}\delta_{21}(\mathbf{x})\ \underbrace{\delta_{\mathrm{LAE}}(\mathbf{x})}_{=-\mathbf{1}} + \sum_{\mathbf{x}_{\mathrm{HII}}}\underbrace{\delta_{21}(\mathbf{x})}_{\simeq\mathbf{0}}\ \delta_{\mathrm{LAE}}(\mathbf{x})\right]\\ &= -\frac{1}{N_{\mathrm{bins}}}\sum_{\mathbf{x}_{\mathrm{HI}}}\left(1-\frac{T_{\mathrm{CMB}}}{T_s(\mathbf{x})}\right)\ x_{\mathrm{HI}}(\mathbf{x})\ (1+\delta_{\mathrm{g}}(\mathbf{x}))\\ &\simeq -\langle\chi_{\mathrm{HI}}\rangle\ \left\langle\left(1-\frac{T_{\mathrm{CMB}}}{T_s}\right)\ (1+\delta_{\mathrm{g}})\right\rangle_{\mathrm{HI}}. &(12.29)\end{aligned}$$

Assuming the post-heating regime $T_{\mathrm{s}} \gg T_{\mathrm{CMB}}$ to be valid in neutral patches, the 21cm-LAE cross correlation at very small scales becomes

$$\xi_{21,\mathrm{LAE}}(r=0) \simeq -\langle x_{\mathrm{HI}}\rangle\ \langle 1+\delta\rangle_{\mathrm{HI}}. \tag{12.30}$$

The small-scale cross-correlation thus directly depends on the progress of reionization via x_{HI}, and the average gas density and thus matter fluctuations in neutral regions. This limit also applies for any representation of δ_{21} that solely shifts the zero-point, e.g. $\delta_{21}(\mathbf{x}) = (\delta T_b(\mathbf{x}) - \langle\delta T_b\rangle)/T_0$.

The cross correlation amplitude profile around LAEs
We now derive the 21cm-LAE cross correlation profile as a function of distance r. Here we limit our calculations to the post-heating regime of the EoR. Separating the 21cm–LAE cross correlation functions into pixels, or

positions, containing ($\delta_{\rm LAE} > -1$) and devoid ($\delta_{\rm LAE} = -1$) of LAEs, we get

$$\begin{aligned}\xi_{21,\rm LAE}(r) &= \frac{1}{N_{\rm bins}} \sum_{n=0}^{N_{\rm bins}} \delta_{21}(\mathbf{x}+\mathbf{r})\ \delta_{\rm LAE}(\mathbf{x}) \\ &= \frac{1}{N_{\rm bins}} \sum_{n=0}^{N_{\rm bins}-N_{\rm LAE}} -\delta_{21}(\mathbf{x}+\mathbf{r})|_{\mathbf{x}\neq\mathbf{x}_{\rm LAE}} \\ &+ \frac{1}{N_{\rm bins}} \sum_{n=0}^{N_{\rm LAE}} \frac{N_{\rm bins}}{N_{\rm LAE}} \delta_{21}(\mathbf{x}+\mathbf{r})|_{\mathbf{x}=\mathbf{x}_{\rm LAE}} \\ &\simeq -\langle\delta_{21}(\mathbf{x})\rangle_{\mathbf{x}} + \frac{1}{N_{\rm LAE}} \sum_{n=0}^{N_{\rm LAE}} \delta_{21}(\mathbf{x}+\mathbf{r})|_{\mathbf{x}=\mathbf{x}_{\rm LAE}} \\ &= -\langle\delta_{21}\rangle + \langle\delta_{21}\rangle^{\rm LAE}(r)\end{aligned}$$

Here $\langle\delta_{21}\rangle^{\rm LAE}$ is the average 21cm signal profile around LAEs, while $\langle\delta_{21}\rangle$ is the average overall 21cm signal, for $N_{\rm bins}$ total number of bins. Correspondingly, $x_{\rm HI}^{\rm LAE}(r)$ and $(1+\delta^{\rm LAE})(r)$ are the average neutral fraction and density profiles around LAEs. We rewrite the previous equation to get

$$\xi_{21,\rm LAE}(r) \quad = \quad -\langle x_{\rm HI}\rangle\langle 1+\delta\rangle_{\rm HI}\left[1 - \frac{\chi_{\rm HI}^{\rm LAE}(r)}{\langle x_{\rm HI}\rangle}\frac{(1+\delta^{\rm LAE})(r)}{\langle 1+\delta\rangle_{\rm HI}}\right]. \tag{12.31}$$

Key factor determining $\xi_{21,\rm LAE}(r)$ are the average neutral hydrogen profile around LAEs, $x_{\rm HI}^{\rm LAE}(r)$ and the average density profile around LAEs. While $x_{\rm HI}^{\rm LAE}(r)$ is determined by the sizes of the ionized regions around LAEs at small r values, it converges to the average neutral hydrogen fraction $\langle x_{\rm HI}\rangle$ as r increases beyond the typical sizes of the ionized regions around LAEs.

12.4 The 21cm signal post-reionization

We have seen in the previous section that during the Epoch of Reionization the 21cm signal of neutral hydrogen and galaxies that trace peaks in the matter density fields (corresponding to halos of a certain mass range) tend to be anti-correlated. This holds if reionization has progressed sufficiently so that high density regions are already ionized (in the so-called 'inside-out' scenario where regions of highest density ionize first). After the completion of reionization the neutral hydrogen is confined to high-density self-shielded system that are protected against radiative ionizing backgrounds. They now are discret sources and tend to positively correlate with the underlying dark matter density. Just as for galaxy redshift surveys, the 21cm linear power spectrum can then be a considered a biased tracer such that

$$P_{21}(z,\,k,\,\mu) = \bar{T}_{21}^2(z)\left[(b_{\rm HI}(z) + f(z)\,\mu^2)^2 P_{\rm m}(z,k) + P_{\rm SN}\right], \tag{12.32}$$

where $\bar{T}_{21}(z)$ is the mean 21cm brightness offset temperature, $b_{\rm HI}(z)$ is the HI linear bias, $f(z)$ the linear growth rate, $\mu = \hat{k}\cdot\hat{z}$, $P_{\rm m}(z,k)$ the total matter power spectrum and $P_{\rm SN}$ is the HI shot noise (Poisson noise) due to the discreteness of HI clouds.

The factor $(1 + f(z)/b_{\rm HI})$ in Eq. 12.32 accounts for redshift space distortions (RSD). RSDs come from the fact that the real-space position of a source in the radial direction is modified by peculiar velocities due to local overdensities. The standard analysis of RSD nowadays focuses on galaxy redshift surveys and makes use of this so-called Kaiser formalism [98]. It relies on several simplifying assumptions, including the linear regime and the distant observer approximation. While the matter power spectrum and the growth function $f(z)$ are determined by our cosmological model, the linear HI bias $b_{\rm HI}$ is constrained via observations and simulations.

For the mean 21cm brightness offset temperature we take the average of Eq. 12.22

$$\bar{T}_{21}\,(z) \approx 27\bar{x}_{\rm HI}\left(\frac{1+z}{10}\frac{0.15}{\Omega_{\rm m,0}h^2}\right)^{1/2}\left(\frac{\Omega_{\rm b,0}h^2}{0.023}\right)\rm mK, \tag{12.33}$$

where again Ω_b and Ω_m are the baryon and total matter density parameters today and $\bar{x}_\mathrm{HI}$ is the mean fraction of neutral hydrogen that in principle still can evolve with redshift after reionization ended. As evolution is much slower post-reionization and due to observational constraints it is often either assumed that $\bar{x}_\mathrm{HI} = \Omega_\mathrm{HI}/\Omega_\mathrm{H}$ is constant with neutral hydrogen matter density $\Omega_\mathrm{H} = 0.74\,\Omega_b$ and ionized hydrogen matter density $\Omega_\mathrm{HI} \sim 10^{-3}$, or slowly evolving as $\Omega_\mathrm{HI} = 10^{-3}\,(1+z)^\beta$ with power-law index β. The linear bias $b_\mathrm{HI}(z)$ is modelled from numerical (hydrodynamical) simulation results [648]. Lastly, the linear growth rate $f(z)$ is model-dependent and is here assumed to only depend on redshift. It can be calculated for models beyond ΛCDM, where generally scale dependence of the growth rate can occur.

12.5 Simulation-based approaches to the 21cm background

As measurements of the 21cm background cover a wide range in time (from the present-day Universe up to reionization and even beyond) and scale (from the globally averaged signal to galaxy-scale fluctuations), so does our toolbox. On the one hand we need approaches that can model the 21cm signal fast over extremely large scales to investigate the Universe's large-scale structure, and on the other hand we need modelling that can access smaller scale information encoded for example in galaxies. In this chapter three common approaches to model the cosmic 21cm background are introduced. These are analytical approaches, semi-numerical simulations, and full numerical simulations that rely on N-body and hydro-dynamical simulations. The three approaches are grouped in ascending order from fast to computationally expensive and from simple to increased realism and complexity of the signal.

12.5.1 Analytical models

Analytical models rely on scaling relations or analytical formulae to describe our signal and relate emission to the underlying dark matter field. As mentioned in section 12.4, since galaxies trace the underlying distribution of dark matter we can relate their emission to dark matter density fluctuations. Fluctuations of line intensity or brightness δI_ν can then be related to dark matter fluctuations δ_m (where we neglect redshift space distortions for now) via

$$\delta I_\nu = \bar{b}_\nu \bar{I}_\nu \delta_\mathrm{m} \tag{12.34}$$

where ν stands for our rest-frame frequency of interest, which is the 21cm line. As dark matter fluctuations at a given scale and redshift can be calculated knowing our cosmological model and initial conditions, we are missing expressions for the mean bias $\bar{b}_\nu$ and mean intensity $\bar{I}_\nu$. Note that in the case of the 21cm background, the Jeans approximation holds well and we thus can easily switch between brightness temperature and intensity. Instead of assuming a value or scaling for $\bar{x}_\mathrm{HI}$ as above in section 12.4, in an approach that divides the matter density field into halos of dark matter (the halo model [649]), we can not only calculate the matter power but also mean intensity via

$$\bar{I}_\nu(z) = \int_{M_\mathrm{min}}^{M_\mathrm{max}} dM \frac{dn}{dM} \frac{L_\mathrm{line}(M,z)}{4\pi D_\mathrm{L}^2} y(z) D_\mathrm{A}^2, \tag{12.35}$$

where M_min and M_max are the minimum and maximum halo masses of halos that host galaxies emitting 21cm radiation, dn/dM is the halo mass function (i.e. the number density of dark matter halos per halo mass M), and D_L and D_A are the luminosity and comoving angular diameter distance, respectively. The factor $y(z) = dr/d\nu = \lambda_\mathrm{line}(1+z)^2/H(z)$, where r is the comoving distance, λ_line is the rest-frame wavelength of an emitted line.

The direct way to estimate the mean intensity of an emission line is to use the observed line luminosity functions (LFs). A common functional form for LFs is the Schechter function [650]

$$\Phi(L)dL = \phi_* \left(\frac{L}{L_*}\right)^\alpha \exp\left(-\frac{L}{L_*}\right) \frac{dL}{L_*}, \tag{12.36}$$

where ϕ_*, α, and L_* are the free parameters obtained by fitting observational data. We can replace the terms dn/dM and dM by $\Phi(L)$ and dL. The observed mean intensity can then be estimated from

$$\bar{I}_\nu(z) = \int_{L_\mathrm{min}}^{L_\mathrm{max}} dL\, \Phi(L) \frac{L}{4\pi D_L^2}\, y(z) D_A^2 \tag{12.37}$$

for a minimum $L_{\rm min}$ and maximum $L_{\rm max}$ luminosity. Lastly, the galaxy bias $\bar{b}_{\rm line}$ is the halo bias $b(M,z)$ [524] weighted by the luminosity of the emitted line, and is given by

$$\bar{b}_{\rm line}(z) = \frac{\int_{M_{\rm min}}^{M_{\rm max}} dM \frac{dn}{dM} L_{\rm line}\, b(M,z)}{\int_{M_{\rm min}}^{M_{\rm max}} dM \frac{dn}{dM} L_{\rm line}}. \tag{12.38}$$

Both halo bias $b(M,z)$ and halo mass function dn/dM can be derived analytically or numerically depending on the cosmology assumed. See arXiv:2303.08752 for a comprehensive review on the halo model in cosmology.

Having estimated the line luminosity, or temperature, in the way described here, we can estimate *statistical* properties of the 21cm brightness offset fluctuations field, such as its power spectrum. While quite non-expensive to compute, analytical models cannot predict certain representations such as imaging of the 21cm signal. Lastly, the relations discussed here apply to times when reionization has finished. During reionization, we need to take into account the gas evolution and ionization states across different non-galactic scales, which do not necessarily positively correlate with the dark matter field as discussed in previous chapters.

12.5.2 Semi-numerical simulations

Semi-numerical simulations are able to create 2D/3D (imaging/tomography) representations of 21cm fluctuation fields. They can do so faster, meaning computationally faster, than numerical hydrodynamical simulations. This allows us to both create larger volumes and explore more possible models. Semi-numerical simulations rely on the assumption of first and second order Lagrangian perturbations for the dark matter density field.

To calculate the 21cm offset brightness temperature, we need to simulate a) the dark matter density field, b) the ionizaton field, or ionized fraction of neutral hydrogen, c) the spin temperature field that depends not only on gas densities and ionization state, but also on gas temperatures and intensities of background radiation. If we work in the post-heating regime $T_{\rm S} >> T_\gamma$, we can restrict our calculation to fields a) and b).

Evolved density fields can be approximated by moving mass particles according to the velocity field derived in the Zel'dovich approximation that was introduced earlier in Sec. 7.1, starting from a Gaussian random field for initial density fluctuations. In this approximation at linear perturbative order, the density fields evolve in redshift as $\delta(z) = \delta(0)\, D(z)$, where $D(z)$ is the linear growth factor with $D(0) = 1$. The position of each mass particle, or equivalently density voxel in our simulation, is then corrected at each redshift with the linear displacement expected from the velocity field. To obtain dark matter halos from the density field, it can be filtered in an excursion-set approach (see Press-Schechter formalism) assuming spherical or ellipsoidal collapse. In the case of ellipsoidal collapse, a collapse threshold $\delta_{\rm c}(M,z)$ which depends both on redshift and filtering scale is employed. The filtering iteself works as follows. At each point $\mathbf{x}$, starting from the largest scales, the field is smoothed with a filtering function, often a real-space top-hat filter, in order to find the largest scale, or mass, such that the the density at that scale fulfills $\delta(\mathbf{x}, M) > \delta_{\rm c}(M,z)$. Is this condition met, mass and position of a new halo have been found. In this simple approach each new halo is not allowed to overlap with a previous halo. Like for the density field itself, positions of halos found through this procedure are displaced at each redshift with the linear displacement expected in the Zel'dovich approximation.

In addition, peculiar velocities, or redshift space distortions on the 21cm signal caused by these, can easily be simulated in the quasi-linear regime. This can be done by correcting Eq. 12.22 with the derivative of the line-of-sight velocity $\mathrm{d}v_{\rm r}/\mathrm{d}r$. In k-space the velocity field at first order can be written as

$$\mathbf{v}(\mathbf{k},z) = \frac{i\mathbf{k}}{k^2}\dot{D}(z)\delta(\mathbf{k}) \ , \tag{12.39}$$

and so the derivative of the line-of-sight velocity v_r can be written in k-space as

$$\begin{aligned}\frac{dv_r}{dr}(\mathbf{k},z) &= ik_r v_r(\mathbf{k},z) \\ &\approx -\frac{k_r^2}{k^2}\dot{D}(z)\delta_{\rm nl}(\mathbf{k}) \ ,\end{aligned} \tag{12.40}$$

where differentiation is performed in k-space.

The ionization field during the reionization process is generated with an excursion-set approach for identifying ionized HII regions. To do so, we filter the evolved density field at different radii, similar as the procedure

described above for halos. From larger to small scales, progressively smaller filtering scales (again often a top-hat filter) are applied at each point of the density field, until the largest scale is found at which the condition

$$f_{\rm coll}\left(\mathbf{x}, M, z\right) \geq \zeta^{-1} \tag{12.41}$$

holds. Here the collapsed mass fraction $f_{\rm coll}$ designates the fraction of mass collapsed to halos at a scale that corresponds to a mass M. The critical parameter is the ionization efficiency ζ, which parametrizes the strength of the radiative field of ionizing radiation. Physically, it is a combination of the fraction of baryons in stars f_*, the number of ionizing photons produced per stellar mass $N_{\gamma/b}$ and the escape fraction of ionizing photons from galaxies into the IGM $f_{\rm esc}$. It can be parametrised as

$$\zeta = 30 \left(\frac{f_{\rm esc}}{0.1}\right) \left(\frac{f_*}{0.05}\right) \left(\frac{N_{\gamma/b}}{4000}\right) \left(\frac{1.5}{1+n_{\rm rec}}\right). \tag{12.42}$$

In this equation the same ionizing efficiency is assumed for each halo. Generally though for example the escape fraction can have a dependence on halo mass. An alternative filtering approach, which includes spatial variation in the radiation field due to differences in recombinations per volume and radiative feedback, simulates the number of ionizing photons per volume e.g. due to recombinations, and then filters ionized regions by requiring the number of ionizing photons produced in a region is equal or higher than the recombination rate. Investigating the impact of these different filtering assumptions on photon conservation in simlations is an ongoing research topic.

Lastly, at times (before reionization and early on during reionization) when the post-heating approximation $T_{\rm S} \gg T_\gamma$ does not hold, we need to follow the evolution of the spin temperature field. As the Wouthuysen-Field effect couples the spin temperature to the kinetic gas temperature $T_{\rm K}$ (and the effective colour temperature of the UV radiation field $T_{\rm c}$, with $T_{\rm K} \to T_{\rm c}$ in most situations), we need to evolve the gas temperature field. Several processes can contribute to heating, such as shock heating from the formation of structures, UV photons and X-rays. During early times, heating through X-rays is the most important process, and we thus discuss here an example of how X-ray emission is modelled in semi-numerical simulations.

Under the assumption that the emission is proportional to the mass fraction in halos, or collapsed fraction, $f_{\rm coll}$, the X-ray emission rate (in photons $\rm s^{-1}$) can be expressed as

$$\frac{{\rm d}\dot{N}_X}{{\rm d}z} = \zeta_X f_* \Omega_{\rm b} \rho_{\rm crit,0} \left(1+\delta_{\rm nl}^R\right) \frac{{\rm d}V}{{\rm d}z} \frac{{\rm d}f_{\rm coll}}{{\rm d}t}\,, \tag{12.43}$$

where ζ_X is the X-ray efficiency, i.e. the number of X-ray photons per solar mass, f_* the fraction of baryons converted to stars, $\delta_{\rm nl}^R$ the non-linear density at scale R corresponding to the smallest mass sources, and ${\rm d}V$ the comoving volume. This emission rate enters the arrival rate $({\rm d}\phi_X/{\rm d}z)$ of X-ray photons of frequency ν, i.e. the number of photons $\rm s^{-1}$ $\rm Hz^{-1}$ seen at position $(\boldsymbol{x}, z)$. Integrating the evolution of the gas temperature equation (12.49), with X-ray heating as the dominant heating source, gives the total X-ray heating rate per baryon ϵ_X. This heating rate also determines part of the Lyα background, where the rate of X-ray conversion to Lyα is given by

$$\epsilon_{X,\alpha} = \epsilon_X \frac{f_{{\rm Ly}\alpha}}{f_{\rm heat}}\,, \tag{12.44}$$

with fraction $f_{{\rm Ly}\alpha}$ of X-ray energy that goes into Lyα photons, and the fraction $f_{\rm heat}$ of electron energy deposited as heat. This gives for the Lyα flux due to X-ray heating (in photons $\rm cm^{-2}$ $\rm s^{-1}$ $\rm Hz^{-1}$ $\rm sr^{-1}$) at position $(\boldsymbol{x}, z)$

$$J_{\alpha,X}\left(\boldsymbol{x}, z\right) = \frac{c}{4\pi H \nu_\alpha} \frac{\epsilon_{X,\alpha}}{h_{\rm Pl}\nu_\alpha}\,. \tag{12.45}$$

The second dominant component of the Lyα background is direct stellar emission, that redshifts into Lyman-n resonance, given by

$$J_{\alpha,*}\left(\boldsymbol{x}, z\right) = \sum_{n=2}^{n_{\rm max}} f_{\rm recycle}\left(n\right) \int_z^{z_{\rm max}} {\rm d}z' \frac{1}{16\pi^2 r_p^2} \frac{{\rm d}\phi_*}{{\rm d}z'}\,, \tag{12.46}$$

with stellar emissivity ($\mathrm{d}\phi_*/\mathrm{d}z'$) and proper separation r_p between z and z'. The stellar emissivity (photons s^{-1} Hz^{-1}) can, similar to equation (12.43) for the the X-ray emissivity, be written as

$$\frac{\mathrm{d}\phi_*}{\mathrm{d}z} = \epsilon\left(\nu_\mathrm{n}\right) f_* \bar{n}_{\mathrm{b},0} \left(1+\delta_\mathrm{nl}^R\right) \frac{\mathrm{d}V}{\mathrm{d}z} \frac{\mathrm{d}f_\mathrm{coll}}{\mathrm{d}t}\,, \tag{12.47}$$

with the number $\epsilon\left(\nu_\mathrm{n}\right)$ of photons per Hz per stellar baryon, for rest frame frequency ν_n at redshift of emission. In this exemplary model, the Lyα background is produced by X-ray heating and stellar emission, while neglecting other possible sources such as quasars or annihilation of dark matter candidates.

The evolution itself of the kinetic gas temperature T_K depends on the local heating history and is coupled to the evolution of the ionized fraction of free electrons x_e of the predominantly neutral regions. The evolution equations can be written as

$$\frac{\mathrm{d}x_\mathrm{e}\left(\mathbf{x},z\right)}{\mathrm{d}z} = \frac{\mathrm{d}t}{\mathrm{d}z}\left[\Lambda_\mathrm{ion} - \alpha_\mathrm{A} C x_\mathrm{e}^2 n_\mathrm{b} f_\mathrm{H}\right]\,, \tag{12.48}$$

$$\frac{\mathrm{d}T_\mathrm{K}\left(\mathbf{x},z\right)}{\mathrm{d}z} = \frac{2}{3k_\mathrm{B}\left(1+x_\mathrm{e}\right)} \frac{\mathrm{d}t}{\mathrm{d}z} \sum_\mathrm{p} \epsilon_\mathrm{p} + \frac{2T_\mathrm{K}}{3n_\mathrm{b}} \frac{\mathrm{d}n_\mathrm{b}}{\mathrm{d}z} - \frac{T_\mathrm{K}}{1+x_\mathrm{e}} \frac{\mathrm{d}x_\mathrm{e}}{\mathrm{d}z}\,, \tag{12.49}$$

with total baryonic number density $n_\mathrm{b} = n_{\mathrm{b},0}\left(1+z\right)^3\left[1+\delta_\mathrm{nl}\right]$, the heating rate ϵ_p of process p per baryon, ionization rate per baryon Λ_ion, case-A recombination coefficient α_A, clumping factor C (that depends on the simulation cell size) and hydrogen number fraction f_H.

12.5.3 Hydrodynamical simulations

In a nutshell, hydrodynamic simulations go beyond dark-matter simulations such as N-body codes that only include the effect of gravity on matter. They also include a fully numerical description for the treatment of gas. For cosmological simulations of structure formation and thus also reionization, mostly two approaches for such a description have been used. Firstly, smoothed particle hydrodynamics (SPH), where the mass of a gaseous fluid is discretised into a number of particles. These particles then move under the combined forces of gravity and hydrodynamics. Secondly, Eulerian or mesh-based methods such as adaptive mesh refinement (AMR), where space is divided into a grid. Gas flow and evolution are then followed in an Eulerian way between neighbouring grid cells. This tracing of gas flows or particle movement and properties becomes computationally expensive for larger volumes and resolved structure formation. Therefore, few models can be explored and volumes that can be simulated are limited. For effects that are computationally too expensive to resolve, especially when aiming to model large-scale effects of the intergalactic medium such as reionization, sub-grid modelling for galaxy formation and evolution such as the typical energy feedback from supernovae needs to be included as well.

If besides the hydrodynamical evolution we need to include the effect of varying ionizing backgrounds as is the case for the modelling of the cosmic 21cm background, we need to include radiative transfer. In simulations that include radiative transfer, rays of radiation are traced for each point through the simulation. Thus the effects of diffuse ionizing radiation propagating through a given inhomogeneous cosmological density field, dependent on on the physical conditions of the gas, can be included. As for each point in the simulation we now need to trace a multitude of line-of-sights through the simulation volume, this surpasses hydrodynamical simulations alone in terms of computational cost. At the same time, hydrodynamical simulations together with radiative transfer are the most 'realistic', or detailed, look one can take at the reionization epoch with computers.

Chapter 13

Cosmological Inference with Machine Learning

Cosmological backgrounds such as the Cosmic Microwave Background or the 21cm background of neutral hydrogen are valuable tools to constrain large-scale structure. To derive (cosmological) model parameters from these mappings, we usually rely on suitable summary statistics such as power spectra. This chapter takes a novel look at cosmological inference using machine learning techniques such as deep learning models for simulation-based and likelihood-free inference.

Quick summary

- Likelihood-free inference relies on a forward model that can simulate data, and a model that can use this information to approximate the posterior.
- This type of inference is useful when the exact full likelihood is not known and/or cannot be computed. This is often the case when our (cosmological) signal is not Gaussian.
- An early example is Approximate Bayesian Computing, where the difference between observed and simulated data is quantified.
- Deep neural networks are models that can be trained with forward-modelled data and corresponding parameter sets to approximate posterior likelihoods.
- Depending on the type of deep learning model, these can learn posteriors for example by variational inference, direct density estimation or inversion.

13.1 Likelihood-free Inference

In cosmology, but also other fields in science, for a given observation the exact full likelihood is typically intractable. Usually, cosmologist rely on writing down a 'good enough', meaning close enough to the true form, likelihood that captures expected theoretical and observational effects. Alternatively, we can strive to take 'complex enough' models to estimate this likelihood from data (observed and synthetic) that need not be analytical. This likelihood 'approximators' are implicit insofar as no likelihood form is typically written down. They always rely on comparing either a learned expectation or simulated data to real data in one form or another. Two common approaches exist for this comparison. In the first one, density estimation methods are used to approximate the likelihood. One method that has become popular very recently are neural density estimators based on neural networks. Here, neural density estimators are for example trained (fitted) with simulated data to represent the desired probability density. Given a set of points $\{x\} \sim p(x)$, the goal is to estimate the corresponding probability density $p(x)$. Secondly, methods such as approximate Bayesian computation (ABC) compare observed and simulated data sets based on some distance measure between the two sets and thus offer a quite direct route to simulation-based inference. But why is this useful?

For example, radio interferometric measurements such as the mapping of the 21cm background of neutral hydrogen have to deal with high levels of foregrounds and noise. Already intrinsically the cosmological fluctuation signal we aim to measure is non-Gaussian, meaning, power spectra as summary statistics do not encode all potentially interesting physical information present in the signal. Other summary statistics and measures such as bispectra or cross-correlations can partly detect this, but lack the flexibility for example of deep learning models. These modern machine learning methods offer the possibility to extract and discard highly non-Gaussian information. In the end, they can also be used to derive cosmological parameters directly from maps. Advantages of modern inference methods rooted in machine learning include the possibility to explore high-dimensional parameter spaces fast, and the fact that inferred parameter values are not biased by the chosen summary statistics ('let the network choose') or likelihood form. An important final word of warning, the prediction capabilities of machine learning methods for inference often rely on the availability of simulations. These need to be computationally efficient to generate large samples of synthetic data, while being as close enough as possible to the observed signal of interest.

13.2 An example: Approximate Bayesian Computing

Until recently before the wider use of deep learning, the probably most well-known method for likelihood-free inference was Approximate Bayesian Computation (ABC), with applications for example to parameter estimation in supernovae cosmology. [651] It is employed in parameter inference in cases where the true likelihood function is intractable, be it because it is unknown or computationally too expensive to be computed. The ABC method relies on the ability to forward simulate synthetic data based on a set of model parameters and the comparison between observed and mock (simulated) data.

Bayesian inference is based on prior distributions of model parameters and the definition of a likelihood for a new set of data drawn from the underlying model. Its goal is to find the posterior distribution of parameters given (observed) data. The posterior likelihood can be calculated either analytically or be sampled for example with Markov Chain Monte Carlo (MCMC [652]). An analytical description for the likelihood function is not always possible. A common solution is to assume a general form for the likelihood function (e.g. Gaussian) followed by MCMC sampling, with the expectation that the assumption on the likelihood is not too far from the true shape. Precondition of the ABC approach is the capability to closely imitate observed data with fast and reliable simulations. To explore the likelihood in this type of inference based on forward simulation we need to quantify the distance between simulated and observed sets of data. The closer a simulation for certain model parameters reproduces observations, the closer it is to the "true" underlying model. A range of different algorithms were developed to optimise the sampling of the parameter space as well as the appropriate distance measure used to compare observations and synthetic data.

The following ingredients are thus indispensable for an ABC algorithm

- prior distributions for input parameters $p(\theta)$ (Bayesian priors),
- a forward model of synthetic data (the simulator),
- a distance function $\rho\,(\mathrm{d}_1, \mathrm{d}_2)$ to compare data set d_1 and d_1.

An ABC algroithm then consists of three main steps:

1. Randomly draw parameter values θ_i from the prior $p(\theta)$ and generate the corresponding synthetic data (simulations) d_i.
2. Calculate the distance between the synthetic data d_i and the observed data d, $\rho_\mathrm{i}\,(\mathrm{d}, \mathrm{d}_\mathrm{i})$.
3. Approximate the posterior $L(\theta; \mathbf{d})$ with the fraction of θ_i that has respective smallest distances.

The choice of the distance function is a crucial step in the design of the ABC algorithm. Possible choices include the Kullback-Leibler (KL) divergence and the Wasserstein distance, which both are also widely used in neural network training.

Let us remind you of Bayes' Theorem and its use to determine the probability of a set of model parameters θ given the data $\mathbf{d}$, or

$$L(\theta; \mathbf{d}) = \frac{\mathbf{f}(\mathbf{d}; \theta)\mathbf{p}(\theta)}{\mathbf{g}(\mathbf{d})}\,. \tag{13.1}$$

As before, $p(\theta)$ is the prior probability, $g(\mathbf{d})$ is the PDF of the data, and $L(\theta;\mathbf{d})$ is the posterior likelihood. Formally, in ABC the likelihood can be written as

$$\mathrm{f(d;\theta)} = \int \mathbf{K}_\epsilon \left[\rho\left(\mathrm{d}, \mathrm{d_S}(\theta_\mathrm{i})\right)\right] \mathrm{f(d;\theta)d\theta} \tag{13.2}$$

where again $\rho\left(\mathrm{d}, \mathrm{d_S}(\theta_\mathrm{i})\right)$ measures the distance between observed and synthetic data, and the kernel K_ϵ allocates smaller weight to higher distance ρ for a chosen bandwidth, or tolerance, ϵ of the kernel. The likelihood can then be approximated as a drawing of n synthetic sets of data with parameter values $\theta_\mathbf{i}$ as

$$\hat{\mathrm{f}}(\mathrm{d};\theta) \approx \frac{1}{\mathrm{n}} \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{K}_\epsilon \left[\rho\left(\mathrm{d}, \mathrm{d_S}(\theta_\mathrm{i})\right)\right] . \tag{13.3}$$

We know from Bayes' Theorem that posterior and likelihood times prior are proportional to each other. In a last step, the approximate posterior is then sampled from the approximate likelihood weighted with the prior. Note that in the limit $\epsilon \to 0$, inference with ABC becomes exact, but for continuous, non-discrete, data the acceptance probability becomes zero. In practice, small values of ϵ require unfeasibly many simulations. For large ϵ, sample efficiency is increased at the expense of inference quality. We thus in practice have to look for a balance between the quality of the likelihood approximation and computational speed.

13.3 Deep learning for inference

13.3.1 A very quick guide to the basics of Deep Learning

This chapter introduces basic terminology and concepts of deep learning models. Goal is not to give a complete introduction to deep learning. This chapter instead strives to give enough information to follow the examples of deep learning models in the following, which will showcase the use of neural networks as a tool for cosmological inference.

Generally, artificial neural networks are build of connected units. They are inspired by the way biological neurons operate in brains. For example, the work of David Hubel and Torsten Wiesel on the hierarchical processing of information in the visual cortex (for which they received the nobel price of medicine in 1981) inspired the neocognitron developed by Kunihiko Fukushima in 1980. [653] The neocognitron is a hierarchical multi-layered artificial neural network that laid the groundwork for convolutional neural networks important today for image recognition and scientific analysis tasks. In their simplest form, the connected units that form neural networks are neurons, or *nodes*, that each take (usually multiple) input signals, process these, and pass them on. Such a node performs a simple matrix multiplication on an input vector b_j to derive an output vector a_i

$$z_\mathrm{i} = \Sigma_\mathrm{j} w_\mathrm{ij} b_\mathrm{j} + c_\mathrm{i} \,, \tag{13.4}$$

where w_ij is the weight for the connection between input element j and output element i and c_i is the bias added to output element i. Both weight and bias for each node are learned, or fitted, parameters of the model. The weight increases or decreases the strength of the input signal, while the bias adds an extra offset. Since this operation alone would be only linear, consecutive linear operations would effectively collapse into one single linear operation. For instance, when using two consecutive steps with the operations $a \times B + c = a'$ and $a' \times B' + c' = a''$, the operation can be rewritten by defining a matrix $D = B \times B'$ and a bias vector $d = c \times B' + c'$, collapsing the two layers into a single linear operation $a \times D + d = a''$. Before a node outputs its value thus a non-linear *activation function* is applied that adds non-linearity to the network model. The choice of activation function(s) is a crucial hyperparameter for network design. For each node the operation thus reads

$$a_\mathrm{i} = f_\mathrm{i}(z_\mathrm{i}) = f_\mathrm{i}\left(\Sigma_\mathrm{j} w_\mathrm{ij} b_\mathrm{j} + c_\mathrm{i}\right) \tag{13.5}$$

so that the activation function f_i is applied. Depending on use-case and data structure, a toolbox of common activation functions exists. One example is the sigmoid defined by $f(x) = \frac{1}{1+e^{-x}}$. It returns 0.5 for $x = 0$, converges to one for high x and converges to zero for high negative values of x. The sigmoid function is often applied to the output layer of classification problems. Another widely used example is the tanh activation

function given by $f(x) = \tanh(x)$ which returns values between -1 and 1. The Rectified Linear Units (ReLu) activation function is one of the more popular activation functions. It is defined by $f(x) = \max(0, x)$. On an intuitive level, the ReLu function might allow to discard output from certain network regions by setting it to negative values.

Typically nodes are aggregated into *layers* to form network architectures. Here an input signal enters a layer, is processed by every node of the layer, and then passed on to the next layer until the output layer is reached. The layers in between input and output are so-called hidden layers and in their presence one typically talks about a deep neural network. An important type of network layers are dense, also called fully connected, layers. Here in each hidden layer each neuron is connected to each single neuron in the layer before and after. Dense layers are for example very much suited to regress parameters and provide an output vector of a given length. Another very common type of layer that is a building block for many neural network architectures are convolutional layers used in the before-mentioned convolutional neural networks. Instead of the matrix multiplication that dense nodes use, they apply convolution kernels, also called filters, whose shape is learned during training of the model. They can be regarded as N-dimensional tensors of weights. The kernel size of the filters are hyperparameters of the architecture. The output of these convolutions are known as feature maps, which are shift invariant. The more filters per layer, the more complex the information that can be captured. Very similar to assembly of information in biological visual processing, layers of convolutions are able to pinpoint hierarchical patterns in data and assemble notions of increasing complexity, going from simple shapes and patterns such as lines to complex representations of objects (such as the famous ability of networks to distinguish objects in images, the 'is this a cat or a dog?').

In the end, neural networks are simply a model like any other that take a set of values as an input, apply some mathematical operations and derive output values or predictions. What makes them special is the hierarchical nature of how information is processed and condensed between different layers, their complexity and flexibility, as well as their ability to improve performance when confronted with increasing amounts of data. In a supervised setting, the free model parameters of a neural network are fitted, a process called training, by guidance of a labelled dataset where true values are known. Network models can easily reach tens of thousands of parameters up to millions or even billions for the largest networks. As a basic rule of thumb, the more the model parameters, the more data we need for fitting, also applies to neural network models. The labels themselves can be created by humans or machines who inspect the data, or be known a priori in case of simulated data. The design choices of the network architecture, like the type of layers, the number of layers as well as details of the training process, are hyperparameters that need in addition be optimised for a given task.

For training, network models are required to be differentiable. They loop over several epochs, where in each epoch predictions are derived with a forward-pass through the network. The quality of predictions is then quantified with a loss, or cost, function L that compares predictions to the 'true' labels of the training dataset. The network parameters themselves are optimised with gradient descent, where for fixed input x and output y the network parameters (weights) can vary. Summarising the weight matrix per layer as $W^l = w^l_{ij}$ the loss between output y and prediction $\hat{y}$ reads

$$L(y, \hat{y}) = L(y, f^l(W^l f^{l-1}(W^{l-1} f^{l-2}(...(W^2 f^1(W^1 x)))))) \,. \tag{13.6}$$

In each layer l the gradients $\partial L / \partial w^l_{ij}$ for each weight with input j and i can then be calculated. Backpropagation efficiently computes the weight gradients at each layer by passing backwards from output to input through the network. At each step, gradients are calculated following the chain rule of differentiation. Starting from the output layer we get

$$\frac{\partial L}{\partial W^l} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial W^l} \,. \tag{13.7}$$

For layer $\mathrm{l} - 1$ we get

$$\frac{\partial L}{\partial W^{l-1}} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial a^{l-1}} \frac{\partial a^{l-1}}{\partial W^{l-1}} . \tag{13.8}$$

For layer $\mathrm{l} - 2$ this reads

$$\frac{\partial L}{\partial W^{l-2}} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial a^{l-1}} \frac{\partial a^{l-1}}{\partial a^{l-2}} \frac{\partial a^{l-2}}{\partial W^{l-2}} \tag{13.9}$$

and so on. There thus will be terms shared between subsequent derivatives. Finally, in this optimisation loop, weight (network parameter) values are then each epoch updated depending on their gradient value. To do so, different optimiser algorithms for networks exist. The simplest solution is stochastic gradient descent where weights are updated according to $w \to w - \eta \nabla L$ for a chosen learning rate η.

To give some examples, the loss function L for regression problems can be as simple as the mean squared error (MSE)

$$L = \frac{1}{n} \sum_{\mathrm{i}=1}^{\mathrm{n}} \left(v_{\mathrm{i,true}} - v_{\mathrm{i,pred}} \right)^2 . \tag{13.10}$$

Here and in the following equation n is the number of data such as images times the total pixel number of output values, $v_{\mathrm{i,true}}$ is the value of the label for image i and $v_{\mathrm{i,pred}}$ the corresponding prediction. The MSE works well with a Gaussian distribution of labels and has a high preference on reducing outliers. Another common loss function for regression tasks is the mean absolute error (MAE)

$$L = \frac{1}{n} \sum_{\mathrm{i}=1}^{\mathrm{n}} |v_{\mathrm{i,true}} - v_{\mathrm{i,pred}}|. \tag{13.11}$$

The MAE works better if the dataset is noisy and has more outliers than a Gaussian distribution.

In practice during training, data is is divided into a training, validation and test set. After each epoch, the neural network is applied on the validation set to compare the accuracy of the predictions on the validation set to those of the training set and to make sure that the neural network is not overfitting on the training set. After the training is stopped when the model has sufficiently converged, the network model is used to evaluate the test set in order to judge its performance on previously 'unseen' data.

Finally, given the high number of network model parameters, overfitting is a common problem. When overfitting occurs, network prediction on the data it was trained on are better than on data it has not seen before. Therefore, a set of *anti-overfitting* measures exist. They include max or average pooling layers that split the input into areas of the same size and only keep the maximum or average value from each area. Often dropout is applied to layers, meaning a fraction of the layer input is set to zero. Also batch normalisation is possible that normalises the layer input to a Gaussian distribution with zero mean and unit standard deviation. The idea is to reduce influence from outliers and at the same time smooth out the loss function to increase the likelihood of finding a better local minimum.

13.3.2 Examples of Deep Learning Models for inference

Recently, using neural networks for inference tasks has come into focus, given their ability to explore high-dimensional parameter spaces. We here describe neural network models that are able to characterise the error on parameter estimates. We start with the idea of moment density networks that are designed to estimate moments of probability distributions, such as mean and variance as well as higher moments such as skewness and kurtosis. Beyond estimates of moments, in cosmological inference we strive to estimate the full shape of the posterior likelihood, while requiring the ability to update likelihoods given new data. Both Bayesian Neural Networks and invertible networks are examples of likelihood-free or simulation-based inference of likelihoods.

Moment density networks

In practice, posterior estimates often serve to compute moments of the estimated distribution, starting with the (co)variance. Therefore in [654] Moment Networks were introduced to side-step the problem of estimating the full posterior and directly estimate its moments. This concept allows for the use of far simpler neural network architectures and thus for example in training faster convergence of results.

Let us first define a model, or function, $\mathcal{F}(\boldsymbol{x})$ of our data $\boldsymbol{x}$ that minimises a mean-squared loss over the distribution $p(\boldsymbol{x}, \boldsymbol{\theta})$ of possible training datasets $\{\boldsymbol{x}_i\}$ for parameters $\{\boldsymbol{\theta}_i\}$,

$$J_0 = \int ||\boldsymbol{\theta} - \mathcal{F}(\boldsymbol{x})||^2 p(\boldsymbol{x}, \boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{x} \, \mathrm{d}\boldsymbol{\theta} \quad , \tag{13.12}$$

then $\mathcal{F}$ evaluated for the observed data is the mean of the posterior distribution $\mathcal{F}(\boldsymbol{x}_{obs}) = \langle \boldsymbol{\theta} \rangle_{\boldsymbol{\theta}|\boldsymbol{x}_{obs}}$. In our case $\mathcal{F}$ is represented by a neural network model. Therefore, a hierarchy of networks generates further moments

of the posterior distribution. For example, let us take a second network model $\mathcal{G}$ after $\mathcal{F}$ is trained and held fixed. If we find a $\mathcal{G}$ that minimises

$$J_1 = \int ||(\boldsymbol{\theta} - \mathcal{F}_{\rm fixed}(\boldsymbol{x}))^2 - \mathcal{G}(\boldsymbol{x})||^2 p(\boldsymbol{x}, \boldsymbol{\theta}) \, {\rm d}\boldsymbol{x} \, {\rm d}\boldsymbol{\theta} \quad , \tag{13.13}$$

then $\mathcal{G}(\boldsymbol{x}_{obs})$ represents the posterior variances. This hierarchy of network models and moments can easily be continued. In practice, one samples the parameter space from the prior $p(\boldsymbol{\theta})$, and combines the functions $\mathcal{F}$ and $\mathcal{G}$ to output the mean, variance, etc for subsets of the full parameter set. When training the network models, marginalisation over the remaining parameters is implicitly done. This result is exact and independent of the true posterior or prior distribution. Moment Networks thus derive moments of marginal posteriors by construction.

Variational inference with Bayesian neural networks

The idea of Bayesian neural networks is based on the notion that network (model) parameters themselves do not need to be point estimates, but instead they themselves can follow a distribution. By choosing a family of distributions, for example each network weight parameter can follow a Gaussian distribution that is characterised by its mean and variance. Instead of one parameter (the weight) in this case we now have two (mean and variance of a Gaussian distribution for this one weight parameter). This way we can capture uncertainties on parameters stemming both from statistical scatter within the data (*aleatoric*) and from uncertainty on the learner, i.e. the network model uncertainty (*epistemic*). The true posterior is thus approximated by a parameterised probability distribution, which we can use for variational parameter inference. In variational inference, a parameterised probability distribution $q(w|\theta)$ over network parameters w given the parameters θ of a family of distributions (e.g. mean and variance for Gaussians) is thus defined as a variational approximation to the true Bayesian posterior $L(w|\boldsymbol{d})$ given the data $\boldsymbol{d}$. The goal of variational inference is to find the set of parameters θ which most closely approximate the true posterior.

The difference between the variational posterior and the exact posterior is determined using a distance measure that compares distributions. Here this distance is typically measured using the Kullback-Leibler (KL) divergence [655]. If a family of distributions with parameters θ is set, the following optimisation problem is then solved during network training

$$\theta^* = \arg\min_{\theta} {\rm KL}[q(w|\theta)||L(w|d)], \tag{13.14}$$

where d denotes the training data and the loss function $F(d, \theta)$ that we seek to minimise is

$$F(d, \theta) = {\rm KL}[q(w|\theta)||L(w|d)]. \tag{13.15}$$

Using the formula for KL divergence, for distributions a and b,

$${\rm KL}[a(x)||b(x)] = \int a(x) \log \frac{a(x)}{b(x)} \, {\rm d}x, \tag{13.16}$$

one can write for the loss function

$$\begin{aligned} F(d, \theta) &= \int q(w|\theta) \log \frac{q(w|\theta)}{L(w|d)} \, {\rm d}w \\ &= \int q(w|\theta) \log \frac{q(w|\theta)}{p(w)p(d|w)} \, {\rm d}w \\ &= \int q(w|\theta) [\log \frac{q(w|\theta)}{p(w)} {\rm d}w - \int q(w|\theta) \log p(d|w)] \, {\rm d}w \\ &= {\rm KL}[q(w|\theta)|p(w)] - \mathbb{E}[\log p(d|w)]_{q(w|\theta)}. \end{aligned} \tag{13.17}$$

This loss function can now be seen to be composed of two components: the first term depends on the prior over the weights $p(w)$ and penalises complexity, and the second term depends on the data and describes how well the model fits to the data in terms of the likelihood.

Equivalently to minimising the above loss function, often the so-called Evidence Lower Bound (ELBO), or variational free energy, is maximised [656]. It reads

$$\begin{aligned} \mathrm{ELBO}(q) &= \mathbb{E}_{q(w)} \log p(w,d) - \mathbb{E}_{q(w)}[\log q(w)] \\ &= \mathbb{E}_{q(w)} \log p(d|w) + \mathbb{E}_{q(w)} \log p(w) - \mathbb{E}_{q(w)}[\log q(w)] \\ &= \mathbb{E}_{q(w)} \log p(d|w) - \mathrm{KL}(q(w)|p(w)). \end{aligned} \tag{13.18}$$

This function is called ELBO as the log evidence is bounded by it such that: $\log p(d) \geq \mathrm{ELBO}$. In this way, Bayesian inference is reduced by variational inference to an optimisation problem, which can be solved with standard deep learning optimisation algorithms.

In practice, often the so-called *Bayes by Backdrop* is performed [657]. To do so, the loss function is reformulated as an expectation with respect to the variational posterior as

$$F(d,\theta) = \int q(w|\theta)[\log q(w|\theta) - \log p(w) - \log p(d|w)]\,\mathrm{d}w. \tag{13.19}$$

To calculate gradients needed for optimisation, this loss function is made differentiable with the reparameterization trick [658]. It maps a sample obtained for the variational posterior $q(w|\theta)$ to a known probability distribution function $q(\epsilon)$ with a differentiable change of variables, such that

$$q(w|\theta)\,\mathrm{d}w = q(\epsilon)\mathrm{d}\epsilon. \tag{13.20}$$

Now that differentiable samples from the variational posterior can be calculated, by drawing Monte Carlo samples $w^{(i)} \sim q(w^{(i)}|\theta)$, the loss function reads

$$F(d,\theta) \approx \sum_{i=1}^{n} \log q(w^{(i)}|\theta) - \log p(w^{(i)}) - \log p(d|w^{(i)}), \tag{13.21}$$

made differentiable for network model optimisation.

Neural density estimators and likelihood-free inference

Another way to estimate the full posterior likelihood for likelihood-free inference is direct density estimation with (neural) density estimators. See for example [659] for its application to cosmology. A variety of neural density estimators (NDEs) exist, that all are able to learn complex distributions from data and simulations.

For likelihood-free inference with density estimators main steps are:

1. Generate parameter-data pairs (θ, d) by running a simulator at parameter values θ.
2. Fit a density estimator $p(d|\theta; w)$ with parameters w to the simulated parameter-data pairs (θ, d).
3. Evaluate the estimated density for observed data d_0 to obtain the (learned, approximate) likelihood $p(d_0|\theta; w)$.

As often the data d is high-dimensional, in order to ensure convergence of the density estimator, the data is compressed (sometimes as well via a summary neural network) to a set of summaries $d \to t$ of lower dimension, which also reduces the number of simulations required.

Invertible neural networks

Often, the forward process from parameter space to observations is a well-defined function, meaning we can create 'mock' observations with simulations. The inverse problem, on the contrary, tends to be ambiguous: one measurement can map to different parameter sets. This can pose a challenge when the posterior parameter distribution, conditioned on an input measurement, is to be determined for parameter inference. Here we want to introduce a particular class of artificial neural networks that is suitable by construction for this inference task, Invertible Neural Networks (INNs). [660] The simulation-based inference approaches presented before aim at directly solving this ambiguous inverse problem, while INNs are able to learn the inverse solution jointly with the well-defined forward process. After training in order to learn to forward process, conditional INNs [661] can be given a specific measurement conditional on random latent variables for the inverse pass through the network. When sampling these latent variables and inputting them together with the observed data, passing inversely ('backwards') through the network efficiently samples the full likelihood distribution.

Two main properties are important for INNs:

1. They are bijective, i.e. both the mapping from inputs to outputs and the inverse exist and can be computed.

2. Both the forward and backward pass, or mapping, through the network have a tractable Jacobian.

In this case it is possible to explicitly calculate posteriors. In practice this means for our task of cosmological parameter inference:

1. A well-defined forward process from input model parameters θ to output d, $\theta \to d$, is used to simulate (synthetic) data d.

2. The invertible network model f_w is trained, i.e. network parameters w are adjusted, until the difference between true posterior likelihood $L(\theta|d)$ and approximated posterior $q_\mathrm{w}(\theta|d)$ is minimised such that $q_\mathrm{w}(\theta|d) \approx L(\theta|d)$ (for the associated loss function the Jacobian is needed, see below).

3. The inverse function f_w^{-1} maps $d_\mathrm{obs} \to \theta$ for observed data d_obs by evaluating the network 'backwards' for both d_obs and random latent space variables z as input and as output the associated parameter set θ.

4. A multitude of samplings of θ via f_w^{-1} evaluated for random z together with fixed d_obs is used for fast exploration of the posterior.

As mentioned, here additional latent output variables z capture input information lost in the output d. The pairs $[d, z]$ of output and latent variables therefore have to be uniquely mapped by the trained function (network model) $f(\theta) = [d, z]$. When the model $f(\theta)$ is optimised for, also its inverse function $\theta = f^{-1}(d, z) = g(d, z)$ is determined. For the latent variables z the distributions $p(z)$ that we map to in training need to be chosen. Usually, they are taken to be Gaussian. Other distributions are in principle also possible. Finally, the INN infers the approximate posterior $q_\mathrm{w}(\theta|d)$ by the inverse mapping $\theta = g(d, z)$ of the known distribution $p(z)$ to the input space θ, conditional on the output d (for inference tasks the input space is the set of parameters that we would like to infer). Similar as before for Bayesian neural networks, we use the Kullback-Leibler (KL) divergence between the true and the approximated posterior as the loss to be minimised. In this case this means

$$w^* = \arg\min_w \mathbb{E}_{p(d)}[\mathrm{KL}(L(\theta|d)||q_\mathrm{w}(\theta|d))] \tag{13.22}$$

with the expectation $\mathbb{E}$ over all possible data $p(d)$. As the true posterior $L(\theta|d)$ does not depend on network parameters w this can be rewritten as

$$w^* = \arg\min_w \mathbb{E}_{p(d)}[\mathbb{E}_{L(\theta|d)}[-\log q_\mathrm{w}(\theta|d)]]. \tag{13.23}$$

As $f_\mathrm{w}(\theta; d) = z$ we can express the approximate posterior by change of variable for probability distributions to be

$$q_\mathrm{w}(\theta|d) = p(f_\mathrm{w}(\theta; d)) \left| \det \left(\frac{\partial f_\mathrm{w}(\theta; d)}{\partial \theta} \right) \right| \tag{13.24}$$

where we need to compute the Jacobian $J_{f_\mathrm{w}} = \frac{\partial f_\mathrm{w}(\theta;d)}{\partial \theta}$. To obtain the loss function for network optimisation, in practice the expectations above are calculated via Monte Carlo estimate for pairs of simulated data d and corresponding parameter set θ.

Bibliography

[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.

[Amendola et al.(2020)] Amendola, L., Bettoni, D., Pinho, A. M., et al. 2020, Universe, 6, 20. doi:10.3390/universe6020020

[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], *Wilkinson Microwave Anisotropy Probe (WMAP) three year results: 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, Cambridge (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* , General Relativity and Gravitation **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 Univeristy 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] M. F. Sohnius, *Introducing Supersymmetry,* Phys. Rept. **128** (1985), 39.

[136] D. Bailin and A. Love, *Supersymmetric gauge field theory and string theory,* Institute of Physics Publishing (1994).

[137] M. Green, J. H. Schwarz and E. Witten, *Superstring theory,* Cambridge University Press, Cambridge (1987).

[138] M. T. Grisaru, W. Siegel and M. Rocek, *Improved methods for supergraphs,* Nucl. Phys. B **159** (1979), 429.

[139] 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.

[140] M. Dine, R. Rohm, N. Seiberg and E. Witten, *Gluino condensation in superstring models,* Phys. Lett. B **156** (1985), 55.

[141] I. Affleck, M. Dine and N. Seiberg, *Dynamical supersymmetry breaking in supersymmetric QCD,* Nucl. Phys. B **241** (1984), 493.

[142] E. Witten, *Dimensional reduction of superstring Models,* Phys. Lett. B **155** (1985), 151.

[143] S. W. Hawking, *The cosmological constant is probably zero,* Phys. Lett. B **134** (1984), 403.

[144] 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.

[145] S. H. H. Tye and I. Wasserman, *A brane world solution to the cosmological constant problem,* Phys. Rev. Lett. **86** (2001), 1682.

[146] J. Yokoyama, *A cosmological constant from degenerate vacua,* Phys. Rev. Lett. **88** (2002), 151302.

[147] S. Mukohyama and L. Randall, *A dynamical approach to the cosmological constant,* Phys. Rev. Lett. **92** (2004), 211302.

[148] G. L. Kane, M. J. Perry and A. N. Zytkow, *An approach to the cosmological constant problem(s),* Phys. Lett. B **609** (2005), 7.

[149] A. D. Dolgov and F. R. Urban, *Dynamical vacuum energy via adjustment mechanism,* Phys. Rev. D **77** (2008), 083503.

[150] G. W. Gibbons, *Aspects of supergravity theories* , in Supersymmetry, Supergravity, and related topics, eds. F. del Aguila, World Scientific (1985), 346-351.

[151] J. Maldacena and C. Nunez, *Supergravity description of field theories on curved manifolds and a no-go theorem,* Int. J. Mod. Phys. **A16** (2001), 822.

[152] J. Polchinski, *Dirichlet-branes and Ramond-Ramond charges,* Phys. Rev. Lett. **75** (1995), 4724.

[153] P. Candelas and X. de la Ossa, *Moduli space of Calabi-Yau manifolds,* Nucl. Phys. B **355** (1991), 455.

[154] S. Gukov, C. Vafa and E. Witten, *CFT's from Calabi-Yau four-folds,* Nucl. Phys. B **584** (2000), 69.

[155] E. Witten, *Non-perturbative superpotentials in string theory,* Nucl. Phys. B **474** (1996), 343.

[156] S. B. Giddings, S. Kachru and J. Polchinski, *Hierarchies from fluxes in string compactifications,* Phys. Rev. D **66** (2002), 106006.

[157] 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.

[158] B. Carter, *Large number coincidences and the anthropic principle in cosmology,* IAU Symposium 63: Confrontation of cosmological theories with observational data (1974), 291.

[159] J. Barrow and F. Tipler, *The cosmological anthropic principle,* Oxford University Press, Oxford (1988).

[160] S. Weinberg, *Anthropic bound on the cosmological constant,* Phys. Rev. Lett. **59** (1987), 2607.

[161] A. D. Linde, *The New inflationary universe scenario,* The very early universe, ed. G. W. Gibbons, S. W. Hawking and S. Siklos, Cambridge University Press (1983), 205.

[162] 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.

[163] A. J. Albrecht and P. J. Steinhardt, *Cosmology for Grand Unified Theories with radiatively induced symmetry breaking,* Phys. Rev. Lett. **48** (1982), 1220.

[164] A. D. Linde, *The inflationary universe,* Rept. Prog. Phys. **47** (1984), 925.

[165] J. D. Brown and C. Teitelboim, *Dynamical neutralization of the cosmological constant,* Phys. Lett. B **195** (1987), 177.

[166] J. D. Brown and C. Teitelboim, *Neutralization of the cosmological constant by membrane creation,* Nucl. Phys. B **297** (1988), 787.

[167] L. F. Abbott, *A mechanism for reducing the value of the cosmological constant,* Phys. Lett. B **150** (1985), 427.

[168] R. Bousso and J. Polchinski, *Quantization of four-form fluxes and dynamical neutralization of the cosmological constant,* JHEP **0006** (2000), 006.

[169] F. Denef and M. R. Douglas, *Distributions of flux vacua,* JHEP **0405** (2004), 072.

[170] R. Blumenhagen, F. Gmeiner, G. Honecker, D. Lust and T. Weigand, *The statistics of supersymmetric D-brane models,* Nucl. Phys. B **713** (2005), 83.

[171] J. Garriga and A. Vilenkin, *A prescription for probabilities in eternal inflation,* Phys. Rev. D **64** (2001), 023507.

[172] R. Bousso, *Holographic probabilities in eternal inflation,* Phys. Rev. Lett. **97** (2006), 191302.

[173] A. Linde, *Towards a gauge invariant volume-weighted probability measure for eternal inflation,* JCAP **0706** (2007), 017.

[174] T. Padmanabhan, *Dark energy: Mystery of the Millennium,* AIP Conf. Proc. **861** (2006), 179.

[175] 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.

[176] C. de Rham, S. Hofmann, J. Khoury and A. J. Tolley, *Cascading Gravity and Degravitation,* JCAP **0802** (2008), 011.

[177] N. Afshordi, *Gravitational Aether and the thermodynamic solution to the cosmological constant problem,* arXiv:astro-ph/0807.2639 (2008).

[178] T. Padmanabhan, *Dark energy and gravity,* Gen. Rel. Grav. **40** (2008), 529.

[179] T. Padmanabhan, *Dark energy and its implications for gravity,* arXiv:gr-qc/0807.2356 (2008).

[180] M. Kunz and D. Sapone, *Dark Energy versus Modified Gravity,* Phys. Rev. Lett. **98** (2007), 121301.

[181] R. R. Caldwell and E. V. Linder, *The limits of quintessence,* Phys. Rev. Lett. **95** (2005), 141301.

[182] A. D. Linde, *Chaotic inflation,* Phys. Lett. B **129** (1983), 177.

[183] A. D. Linde, *Inflation and quantum cosmology,* in: Three hundred years of gravitation, Eds.: S. W. Hawking and W. Israel, Cambridge Univeristy Press (1987), 604.

[184] R. Kallosh, J. Kratochvil, A. Linde, E. V. Linder and M. Shmakova, *Observational bounds on cosmic doomsday,* JCAP **0310** (2003), 015.

[185] A. de la Macorra and G. Piccinelli, *General scalar fields as quintessence,* Phys. Rev. D **61** (2000), 123503.

[186] S. C. C. Ng, N. J. Nunes and F. Rosati, *Applications of scalar attractor solutions to cosmology,* Phys. Rev. D **64** (2001), 083510.

[187] T. Barreiro, E. J. Copeland and N. J. Nunes, *Quintessence arising from exponential potentials,* Phys. Rev. D **61** (2000), 127301.

[188] V. Sahni and L. M. Wang, *A new cosmological model of quintessence and dark matter,* Phys. Rev. D **62** (2000), 103517.

[189] A. R. Liddle, A. Mazumdar and F. E. Schunck, *Assisted inflation,* Phys. Rev. D **58** (1998), 061301.

[190] 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.

[191] S. A. Kim, A. R. Liddle and S. Tsujikawa, *Dynamics of assisted quintessence,* Phys. Rev. D **72** (2005), 043506.

[192] J. Ohashi and S. Tsujikawa, *Assisted dark energy,* arXiv:gr-qc/0909.3924 (2009).

[193] P. J. E. Peebles and A. Vilenkin, *Quintessential inflation,* Phys. Rev. D **59** (1999), 063505.

[194] L. Kofman, A. D. Linde and A. A. Starobinsky, *Towards the theory of reheating after inflation,* Phys. Rev. D **56** (1997), 3258.

[195] G. N. Felder, L. Kofman and A. D. Linde, *Instant preheating,* Phys. Rev. D **59** (1999), 123523.

[196] R. Bean, S. H. Hansen and A. Melchiorri, *Early universe constraints on dark energy,* Phys. Rev. D **64** (2001), 103508.

[197] C. J. Copi, D. N. Schramm and M. S. Turner, *Assessing big bang nucleosynthesis,* Phys. Rev. Lett. **75** (1995), 3981.

[198] P. J. Kernan and S. Sarkar, *No crisis for big bang nucleosynthesis,* Phys. Rev. D **54** (1996), 3681.

[199] C. F. Kolda and D. H. Lyth, *Quintessential difficulties,* Phys. Lett. B **458** (1999), 197.

[200] T. R. Taylor, G. Veneziano and S. Yankielowicz, *Supersymmetric QCD and its massless limit: an effective Lagrangian analysis,* Nucl. Phys. B **218** (1983), 493.

[201] E. Witten, *The cosmological constant from the viewpoint of string theory,* arXiv:hep-ph/0002297 (2000).

[202] R. Kallosh, A. D. Linde, S. Prokushkin and M. Shmakova, *Gauged supergravities, de Sitter space and cosmology,* Phys. Rev. D **65** (2002), 105016.

[203] R. Kallosh, A. Linde, S. Prokushkin and M. Shmakova, *Supergravity, dark energy and the fate of the universe,* Phys. Rev. D **66** (2002), 123503.

[204] P. Fre, M. Trigiante and A. Van Proeyen, *Stable de Sitter vacua from $\mathcal{N} = 2$ supergravity,* Class. Quant. Grav. **19** (2002), 4167.

[205] Y. Nomura, T. Watari and T. Yanagida, *Quintessence axion potential induced by electroweak instanton effects,* Phys. Lett. B **484** (2000), 103.

[206] K. Choi, *String or M theory axion as a quintessence,* Phys. Rev. D **62** (2000), 043509.

[207] J. E. Kim and H. P. Nilles, *A quintessential axion,* Phys. Lett. B **553** (2003), 1.

[208] L. J. Hall, Y. Nomura and S. J. Oliver, *Evolving dark energy with w not equal -1,* Phys. Rev. Lett. **95** (2005), 141302.

[209] J. E. Lidsey, D. Wands and E. J. Copeland, *Superstring cosmology,* Phys. Rept. **337** (2000), 343.

[210] M. Gasperini, F. Piazza and G. Veneziano, *Quintessence as a run-away dilaton,* Phys. Rev. D **65** (2002), 023508.

[211] T. Damour, F. Piazza and G. Veneziano, *Runaway dilaton and equivalence principle violations,* Phys. Rev. Lett. **89** (2002), 081601.

[212] G. Veneziano, *Large-N bounds on, and compositeness limit of, gauge and gravitational interactions,* JHEP **0206** (2002), 051.

[213] K. i. Maeda, *Towards the Einstein-Hilbert action via conformal transformation,* Phys. Rev. D **39** (1989), 3159.

[214] A. A. Starobinsky, *How to determine an effective potential for a variable cosmological term,* JETP Lett. **68** (1998), 757.

[215] D. Huterer and M. S. Turner, *Revealing quintessence,* Phys. Rev. D **60** (1999), 081301.

[216] 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.

[217] T. Chiba and T. Nakamura, *Feasibility of reconstructing the quintessential potential using SN Ia data,* Phys. Rev. D **62** (2000), 121301.

[218] 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.

[219] U. Alam, V. Sahni, T. D. Saini and A. A. Starobinsky, *Exploring the expanding universe and dark rnergy using the statefinder diagnostic,* Mon. Not. Roy. Astron. Soc. **344** (2003), 1057.

[220] 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.

[221] G. Efstathiou, *Constraining the equation of state of the universe from distant Type Ia Supernovae and Cosmic Microwave Background Anisotropies,* Mon. Not. R. Astron. Soc. **342** (2000), 810.

[222] M. Chevalier and D. Polarski, *Accelerating universes with scaling dark matter,* Int. J. Mod. Phys. D **10** (2001), 213.

[223] J. Weller and A. J. Albrecht, *Future Supernovae observations as a probe of dark energy,* Phys. Rev. D **65** (2002), 103512.

[224] 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.

[225] 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.

[226] P. S. Corasaniti and E. J. Copeland, *A model independent approach to the dark energy equation of state,* Phys. Rev. D **67** (2003), 063521.

[227] E. V. Linder, *Exploring the expansion history of the universe,* Phys. Rev. Lett. **90** (2003), 091301.

[228] S. Nesseris and L. Perivolaropoulos, *A comparison of cosmological models using recent supernova data,* Phys. Rev. D **70** (2004), 043531.

[229] R. Lazkoz, S. Nesseris and L. Perivolaropoulos, *Exploring Cosmological Expansion Parametrizations with the Gold SnIa Dataset,* JCAP **0511** (2005), 010.

[230] 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.

[231] A. Upadhye, M. Ishak and P. J. Steinhardt, *Dynamical dark energy: Current constraints and forecasts,* Phys. Rev. D **72** (2005), 063501.

[232] R. Crittenden, E. Majerotto and F. Piazza, *Measuring deviations from a cosmological constant: a field-space parameterization,* Phys. Rev. Lett. **98** (2007), 251301.

[233] K. Ichikawa and T. Takahashi, *Dark energy parametrizations and the curvature of the universe,* JCAP **0702** (2007), 001.

[234] 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.

[235] V. Sahni, T. D. Saini, A. A. Starobinsky and U. Alam, *Statefinder – a new geometrical diagnostic of dark energy,* JETP Lett. **77** (2003), 201.

[236] V. Sahni and A. Starobinsky, *Reconstructing dark energy,* Int. J. Mod. Phys. D **15** (2006), 2105.

[237] E. V. Linder, *The paths of quintessence,* Phys. Rev. D **73** (2006), 063010.

[238] C. Armendariz-Picon, T. Damour and V. F. Mukhanov, *k-Inflation,* Phys. Lett. B **458** (1999), 209.

[239] 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.

[240] F. Piazza and S. Tsujikawa, *Dilatonic ghost condensate as dark energy,* JCAP **0407** (2004), 004.

[241] M. R. Garousi, *Tachyon couplings on non-BPS D-branes and Dirac-Born-Infeld action,* Nucl. Phys. B **584** (2000), 284.

[242] A. Sen, *Rolling Tachyon,* JHEP **0204** (2002), 048.

[243] D. Kutasov and V. Niarchos, *Tachyon effective actions in open string theory,* Nucl. Phys. B **666** (2003), 56.

[244] T. Padmanabhan, *Accelerated expansion of the universe driven by tachyonic matter,* Phys. Rev. D **66** (2002), 021301.

[245] L. R. W. Abramo and F. Finelli, *Cosmological dynamics of the tachyon with an inverse power-law potential,* Phys. Lett. B **575** (2003), 165.

[246] J. M. Aguirregabiria and R. Lazkoz, *Tracking solutions in tachyon cosmology,* Phys. Rev. D **69** , 123502 (2004).

[247] 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.

[248] 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.

[249] E. Silverstein and D. Tong, *Scalar speed limits and cosmology: Acceleration from D-cceleration,* Phys. Rev. D **70** (2004), 103505.

[250] M. Alishahiha, E. Silverstein and D. Tong, *DBI in the sky,* Phys. Rev. D **70** (2004), 123505.

[251] J. M. Maldacena, *The large N limit of superconformal field theories and supergravity,* Adv. Theor. Math. Phys. **2** (1998), 231.

[252] J. Martin and M. Yamaguchi, *DBI-essence,* Phys. Rev. D **77** (2008), 123508.

[253] Z. K. Guo and N. Ohta, *Cosmological Evolution of Dirac-Born-Infeld Field,* JCAP **0804** (2008), 035.

[254] J. Garriga and V. F. Mukhanov, *Perturbations in k-inflation,* Phys. Lett. B **458** (1999), 219.

[255] 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.

[256] J. M. Cline, S. Jeon and G. D. Moore, *The phantom menaced: Constraints on low-energy effective ghosts,* Phys. Rev. D **70** (2004), 043543.

[257] E. Babichev, V. Mukhanov and A. Vikman, *k-essence, superluminal propagation, causality and emergent geometry* , JHEP **0802** (2008), 101.

[258] M. Malquarti, E. J. Copeland and A. R. Liddle, *K-essence and the coincidence problem,* Phys. Rev. D **68** (2003), 023512.

[259] C. Bonvin, C. Caprini and R. Durrer, *A no-go theorem for k-essence dark energy,* Phys. Rev. Lett. **97** (2006), 081303.

[260] P. Singh, M. Sami and N. Dadhich, *Cosmological dynamics of phantom field,* Phys. Rev. D **68** (2003), 023522.

[261] M. Sami and A. Toporensky, *Phantom Field and the Fate of universe,* Mod. Phys. Lett. A **19** (2004), 1509.

[262] P. Sreekumar *et al.* [EGRET Collaboration], *EGRET observations of the extragalactic gamma ray emission,* Astrophys. J. **494** (1998), 523.

[263] R. Gannouji, D. Polarski, A. Ranquet and A. A. Starobinsky, *Scalar-tensor models of normal and phantom dark energy,* JCAP **0609** (2006), 016.

[264] B. Feng, X. L. Wang and X. M. Zhang, *Dark energy constraints from the cosmic age and supernova,* Phys. Lett. B **607** (2005), 35.

[265] 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.

[266] 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.

[267] L. Perivolaropoulos, *Reconstruction of extended quintessence potentials from the SnIa gold dataset,* JCAP **0510** (2005), 001.

[268] J. Martin, C. Schimd and J. P. Uzan, *Testing for $w < -1$ in the solar system,* Phys. Rev. Lett. **96** (2006), 061303.

[269] 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.

[270] L. Amendola, *Scaling solutions in general non-minimal coupling theories,* Phys. Rev. D **60** (1999), 043501.

[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 and D. Pavon, *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] 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.

[319] J. Khoury and A. Weltman, *Chameleon fields: Awaiting surprises for tests of gravity in space,* Phys. Rev. Lett. **93** (2004), 171104.

[320] J. Khoury and A. Weltman, *Chameleon cosmology,* Phys. Rev. D **69** (2004), 044026.

[321] T. Tamaki and S. Tsujikawa, *Revisiting chameleon gravity: Thin-shell and no-shell fields with appropriate boundary conditions,* Phys. Rev. D **78** (2008), 084028.

[322] S. Tsujikawa, T. Tamaki and R. Tavakol, *Chameleon scalar fields in relativistic gravitational backgrounds,* JCAP **0905** (2009), 020.

[323] C. M. Will, *The confrontation between general relativity and experiment,* Living Rev. Rel. **9** (2005), 3.

[324] C. D. Hoyle *et al., Sub-millimeter tests of the gravitational inverse-square law,* Phys. Rev. D **70** (2004), 042004.

[325] 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.

[326] D. F. Mota and D. J. Shaw, *Strongly coupled chameleon fields: New horizons in scalar field theory,* Phys. Rev. Lett. **97** (2006), 151102.

[327] 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.

[328] L. Amendola, M. Quartin, S. Tsujikawa and I. Waga, *Challenges for scaling cosmologies,* Phys. Rev. D **74** (2006), 023525.

[329] S. Tsujikawa and M. Sami, *A unified approach to scaling solutions in a general cosmological background,* Phys. Lett. B **603** (2004), 113.

[330] S. Mizuno and K. i. Maeda, *Quintessence in a brane world,* Phys. Rev. D **64** (2001), 123521.

[331] E. J. Copeland, S. J. Lee, J. E. Lidsey and S. Mizuno, *Generalized cosmological scaling solutions,* Phys. Rev. D **71** (2005), 023526.

[332] 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.

[333] R. J. Scherrer, *Purely kinetic k-essence as unified dark matter,* Phys. Rev. Lett. **93** (2004), 011301.

[334] D. Carturan, and F. Finelli, *Cosmological effects of a class of fluid dark energy models,* Phys. Rev. D **68** (2003), 103501.

[335] H. Sandvik, M. Tegmark, M. Zaldarriaga and I. Waga, *The end of unified dark matter?,* Phys. Rev. D **69** (2004), 123524.

[336] P. P. Avelino *et al., The onset of the non-linear regime in unified dark matter models,* Phys. Rev. D **69** (2004), 041301.

[337] L. Amendola, F. Finelli, C. Burigana and D. Carturan, *WMAP and the generalized Chaplygin gas,* JCAP **0307** (2003), 005.

[338] L. Amendola, I. Waga and F. Finelli, *Observational constraints on silent quartessence,* JCAP **0511** (2005), 009.

[339] 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.

[340] D. Bertacca, S. Matarrese and M. Pietroni, *Unified dark matter in scalar field cosmologies,* Mod. Phys. Lett. A **22** (2007), 2893.

[341] T. Fukuyama, M. Morikawa and T. Tatekawa, *Cosmic structures via Bose Einstein condensation and its collapse,* JCAP **0806** (2008), 033.

[342] S. Nojiri, S. D. Odintsov and S. Tsujikawa, *Properties of singularities in (phantom) dark energy universe,* Phys. Rev. D **71** (2005), 063004.

[343] J. D. Barrow, *Sudden future singularities,* Class. Quant. Grav. **21** (2004), L79.

[344] H. Stefancic, *'Expansion' around the vacuum equation of state: Sudden future singularities and asymptotic behavior,* Phys. Rev. D **71** (2005), 084024.

[345] I. H. Brevik and O. Gorbunova, *Dark Energy and Viscous Cosmology,* Gen. Rel. Grav. **37** (2005), 2039.

[346] M. P. Dabrowski, *Statefinders, higher-order energy conditions and sudden future singularities,* Phys. Lett. B **625** (2005), 184.

[347] M. Bouhmadi-Lopez, P. F. Gonzalez-Diaz and P. Martin-Moruno, *Worse than a big rip?,* Phys. Lett. B **659** (2008), 1.

[348] J. D. Barrow, G. Galloway and F. Tipler, *The closed-universe recollapse conjecture* , Mon. Not. Roy. astr. Soc. **223** (1986), 835.

[349] J. D. Barrow, *Graduated inflationary universes,* Phys. Lett. B **235** (1990), 40.

[350] M. Sami, P. Singh and S. Tsujikawa, *Avoidance of future singularities in loop quantum cosmology,* Phys. Rev. D **74** (2006), 043514.

[351] D. Samart and B. Gumjudpai, *Phantom field dynamics in loop quantum cosmology,* Phys. Rev. D **76** (2007), 043514.

[352] M. Bojowald, *Loop quantum cosmology,* Living Rev. Rel. **8** (2005), 11.

[353] A. Ashtekar, T. Pawlowski and P. Singh, *Quantum nature of the big bang: An analytical and numerical investigation,* Phys. Rev. D **73** (2006), 124038.

[354] 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.

[355] T. Chiba, *1/R gravity and scalar-tensor gravity,* Phys. Lett. B **575** (2003), 1.

[356] A. D. Dolgov and M. Kawasaki, *Can modified gravity explain accelerated cosmic expansion?,* Phys. Lett. B **573** , 1 (2003).

[357] 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.

[358] G. J. Olmo, *Post-Newtonian constraints on $f(R)$ cosmologies in metric and Palatini formalism,* Phys. Rev. D **72** (2005), 083505.

[359] V. Faraoni, *Solar system experiments do not yet veto modified gravity models,* Phys. Rev. D **74** (2006), 023529.

[360] I. Navarro and K. Van Acoleyen, *$f(R)$ actions, cosmic acceleration and local tests of gravity,* JCAP **0702** (2007), 022.

[361] 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.

[362] S. M. Carroll, I. Sawicki, A. Silvestri and M. Trodden, *Modified-source gravity and cosmological structure formation,* New J. Phys. **8** (2006), 323.

[363] Y. S. Song, W. Hu and I. Sawicki, *The large scale structure of $f(R)$ gravity,* Phys. Rev. D **75** (2007), 044004.

[364] 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.

[365] 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.

[366] W. Hu and I. Sawicki, *Models of $f(R)$ cosmic acceleration that evade solar-system tests,* Phys. Rev. D **76** (2007), 064004.

[367] A. A. Starobinsky, *Disappearing cosmological constant in $f(R)$ gravity,* JETP Lett. **86** (2007), 157.

[368] S. A. Appleby and R. A. Battye, *Do consistent $f(R)$ models mimic General Relativity plus Λ?,* Phys. Lett. B **654** (2007), 7.

[369] S. Tsujikawa, *Observational signatures of $f(R)$ dark energy models that satisfy cosmological and local gravity constraints,* Phys. Rev. D **77** (2008), 023507.

[370] V. Muller, H. J. Schmidt and A. A. Starobinsky, *The stability of the de-Sitter spacetime in fourth order gravity,* Phys. Lett. B **202** (1988), 198.

[371] V. Faraoni, *de Sitter attractors in generalized gravity,* Phys. Rev. D **70** (2004), 044037.

[372] 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.

[373] 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.

[374] B. Li and J. D. Barrow, *The cosmology of $f(R)$ gravity in the metric variational approach,* Phys. Rev. D **75** (2007), 084010.

[375] R. Gannouji, B. Moraes and D. Polarski, *The growth of matter perturbations in $f(R)$ models,* Phys. Rev. D **0902** (2009), 034.

[376] S. Capozziello and S. Tsujikawa, *Solar system and equivalence principle constraints on $f(R)$ gravity by chameleon approach,* Phys. Rev. D **77** (2008), 107501.

[377] 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.

[378] N. Deruelle, M. Sasaki and Y. Sendouda, *'Detuned' $f(R)$ gravity and dark energy,* Phys. Rev. D **77** (2008), 124024.

[379] 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.

[380] A. V. Frolov, *A singularity problem with $f(R)$ dark energy,* Phys. Rev. Lett. **101** (2008), 061103.

[381] T. Kobayashi and K. i. Maeda, *Relativistic stars in $f(R)$ gravity, and absence thereof,* Phys. Rev. D **78** (2008), 064019.

[382] E. Babichev and D. Langlois, *Relativistic stars in $f(R)$ gravity,* arXiv:gr-qc/0904.1382 (2009).

[383] A. Upadhye and W. Hu, *The existence of relativistic stars in $f(R)$ gravity,* arXiv:astro-ph/0905.4055 (2009).

[384] D. N. Vollick, *Curvature corrections as the source of the cosmological acceleration,* Phys. Rev. D **68** (2003), 063510.

[385] D. N. Vollick, *On the viability of the Palatini form of $1/R$ gravity,* Class. Quant. Grav. **21** (2004), 3813.

[386] X. Meng and P. Wang, *Modified Friedmann equations in R^{-1}-modified gravity,* Class. Quant. Grav. **20** (2003), 4949.

[387] E. E. Flanagan, *The conformal frame freedom in theories of gravitation,* Class. Quant. Grav. **21** (2004), 3817.

[388] T. P. Sotiriou, *Constraining $f(R)$ gravity in the Palatini formalism,* Class. Quant. Grav. **23** (2006), 1253.

[389] T. P. Sotiriou, *$f(R)$ gravity and scalar-tensor theory,* Class. Quant. Grav. **23** (2006), 5117.

[390] 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.

[391] M. Amarzguioui, O. Elgaroy, D. F. Mota and T. Multamaki, *Cosmological constraints on $f(R)$ gravity theories within the Palatini approach,* Astron. Astrophys. **454** (2006), 707.

[392] T. Koivisto and H. Kurki-Suonio, *Cosmological perturbations in the Palatini formulation of modified gravity,* Class. Quant. Grav. **23** (2006), 2355.

[393] 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.

[394] E. E. Flanagan, *Palatini form of $1/R$ gravity,* Phys. Rev. Lett. **92** (2004), 071101.

[395] A. Iglesias, N. Kaloper, A. Padilla and M. Park, *How (Not) to Palatini,* Phys. Rev. D **76** (2007), 104001.

[396] 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.

[397] T. P. Sotiriou and V. Faraoni, *$f(R)$ theories of gravity,* arXiv:gr-qc/0805.1726 (2008).

[398] Y. Fujii and K. Maeda, *The scalar-tensor theory of gravitation,* Cambridge University Press, Cambridge (2003).

[399] G. Esposito-Farese and D. Polarski, *Scalar-tensor gravity in an accelerating universe,* Phys. Rev. D **63** (2001), 063504.

[400] D. F. Torres, *Quintessence, super-quintessence and observable quantities in Brans-Dicke and non-minimally coupled theories,* Phys. Rev. D **66** (2002), 043522.

[401] S. Nesseris and L. Perivolaropoulos, *Evolving Newton's constant, extended gravity theories and SnIa data analysis,* Phys. Rev. D **73** (2006), 103511.

[402] T. Damour and K. Nordtvedt, *Tensor - scalar cosmological models and their relaxation toward general relativity,* Phys. Rev. D **48** (1993), 3436.

[403] J. D. Bekenstein, *Relativistic gravitation theory for the MOND paradigm,* Phys. Rev. D **70** (2004), 083509.

[404] 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.

[405] A. De Felice, M. Hindmarsh and M. Trodden, *Ghosts, instabilities, and superluminal propagation in modified gravity models,* JCAP **0608** (2006), 005.

[406] G. Calcagni, B. de Carlos and A. De Felice, *Ghost conditions for Gauss-Bonnet cosmologies,* Nucl. Phys. B **752** (2006), 404.

[407] A. Nunez and S. Solganik, *Ghost constraints on modified gravity,* Phys. Lett. B **608** (2005), 189.

[408] S. Nojiri, S. D. Odintsov and M. Sasaki, *Gauss-Bonnet dark energy,* Phys. Rev. D **71** (2005), 123509.

[409] G. Calcagni, S. Tsujikawa and M. Sami, *Dark energy and cosmological solutions in second-order string gravity,* Class. Quant. Grav. **22** (2005), 3977.

[410] B. M. N. Carter and I. P. Neupane, *Towards inflation and dark energy cosmologies from modified Gauss-Bonnet theory,* JCAP **0606** (2006), 004.

[411] L. Amendola, C. Charmousis and S. C. Davis, *Constraints on Gauss-Bonnet gravity in dark energy cosmologies,* JCAP **0612** (2006), 020.

[412] T. Koivisto and D. F. Mota, *Cosmology and astrophysical constraints of Gauss-Bonnet dark energy,* Phys. Lett. B **644** (2007), 104.

[413] T. Koivisto and D. F. Mota, *Gauss-Bonnet quintessence: Background evolution, large scale structure and cosmological constraints,* Phys. Rev. D **75** (2007), 023518.

[414] 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.

[415] 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.

[416] 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.

[417] L. Amendola, C. Charmousis and S. C. Davis, *Solar system constraints on Gauss-Bonnet mediated dark energy,* JCAP **0710** (2007), 004.

[418] De Felice, Antonio and Tsujikawa, Shinji, *Conditions for the cosmological viability of the most general scalar-tensor theories and their applications to extended Galileon dark energy models,* JCAP **1202** (2012), 007.

[419] M. Ostrogradsky, *Mémoires sur les équations différentielles, relatives au problème des isopérimètres,* Mem. Acad. St. Petersbourg **6** 4 (1850), 385

[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] L. Randall and R. Sundrum, *A large mass hierarchy from a small extra dimension,* Phys. Rev. Lett. **83** (1999), 3370.

[426] L. Randall and R. Sundrum, *An alternative to compactification,* Phys. Rev. Lett. **83** (1999), 4690.

[427] P. Binetruy, C. Deffayet and D. Langlois, *Non-conventional cosmology from a brane-universe,* Nucl. Phys. B **565** (2000), 269.

[428] P. Binetruy, C. Deffayet, U. Ellwanger and D. Langlois, *Brane cosmological evolution in a bulk with cosmological constant,* Phys. Lett. B **477** (2000), 285.

[429] T. Shiromizu, K. i. Maeda and M. Sasaki, *The Einstein equations on the 3-brane world,* Phys. Rev. D **62** (2000), 024012.

[430] G. R. Dvali, G. Gabadadze and M. Porrati, *4D gravity on a brane in 5D Minkowski space,* Phys. Lett. B **485** (2000), 208.

[431] G. R. Dvali and G. Gabadadze, *Gravity on a brane in infinite-volume extra space,* Phys. Rev. D **63** (2001), 065007.

[432] C. Deffayet, *Cosmology on a brane in Minkowski bulk,* Phys. Lett. B **502** (2001), 199.

[433] C. Deffayet, G. R. Dvali and G. Gabadadze, *Accelerated universe from gravity leaking to extra dimensions,* Phys. Rev. D **65** (2002), 044023.

[434] I. Sawicki and S. M. Carroll, *Cosmological Structure Evolution and CMB Anisotropies in DGP Braneworlds,* arXiv:astro-ph/0510364 (2005).

[435] M. Fairbairn and A. Goobar, *Supernova limits on brane world cosmology,* Phys. Lett. B **642** (2006), 432.

[436] R. Maartens and E. Majerotto, *Observational constraints on self-accelerating cosmology,* Phys. Rev. D **74** (2006), 023004.

[437] U. Alam and V. Sahni, *Confronting braneworld cosmology with supernova data and baryon oscillations,* Phys. Rev. D **73** (2006), 084024.

[438] Y. S. Song, I. Sawicki and W. Hu, *Large-scale tests of the DGP model,* Phys. Rev. D **75** (2007), 064003.

[439] 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.

[440] K. Koyama and R. Maartens, *Structure formation in the DGP cosmological model,* JCAP **0601** (2006), 016.

[441] R. Durrer and R. Maartens, *Dark energy and modified gravity,* arXiv:astro-ph/0811.4132 (2008).

[442] D. Gorbunov, K. Koyama and S. Sibiryakov, *More on ghosts in DGP model,* Phys. Rev. D **73** (2006), 044016.

[443] C. Charmousis, R. Gregory, N. Kaloper and A. Padilla, *DGP specteroscopy,* JHEP **0610** (2006), 066.

[444] K. Tomita, *Distances and lensing in cosmological void models,* Astrophys. J. **529** (2000), 38.

[445] K. Tomita, *A local void and the accelerating universe,* Mon. Not. Roy. Astron. Soc. **326** (2001), 287.

[446] M. N. Celerier, *Do we really see a cosmological constant in the supernovae data?,* Astron. Astrophys. **353** (2000), 63.

[447] 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.

[448] H. Alnes, M. Amarzguioui and O. Gron, *An inhomogeneous alternative to dark energy?,* Phys. Rev. D **73** (2006), 083519.

[449] 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).

[450] 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.

[451] J. Garcia-Bellido and T. Haugboelle, *Confronting Lemaître-Tolman-Bondi models with observational cosmology,* JCAP **0804** (2008), 003.

[452] K. Enqvist, *Lemaître-Tolman-Bondi model and accelerating expansion,* Gen. Rel. Grav. **40** (2008), 451.

[453] J. Kristian and R. K. Sachs, *Observations in cosmology,* Astrophys. J. **143** (1966), 379.

[454] H. Alnes and M. Amarzguioui, *CMB anisotropies seen by an off-center observer in a spherically symmetric inhomogeneous universe,* Phys. Rev. D **74** (2006), 103520.

[455] 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).

[456] I. Masina and A. Notari, *The cold cpot as a large void: Rees-Sciama effect on CMB power spectrum and bispectrum,* JCAP **0902** (2009), 019.

[457] 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.

[458] R. R. Caldwell and A. Stebbins, *A test of the Copernican Principle,* Phys. Rev. Lett. **100** (2008), 191302.

[459] F. Occhionero, P. Santangelo, and N. Vittorio, *Holes in cosmology* , Astron. Astrophys. **117** (1983), 365.

[460] T. Biswas, R. Mansouri and A. Notari, *Non-linear structure formation and apparent acceleration: an Investigation,* JCAP **0712** (2007), 017.

[461] C. Clarkson, B. Bassett and T. C. Lu, *A general test of the Copernican Principle,* Phys. Rev. Lett. **101** (2008), 011301.

[462] M. H. Partovi and B. Mashhoon, *Toward verification of large-scale homogeneity in cosmology* , Astrophys. J. **276** (1984), 4.

[463] C. Quercellini, M. Quartin and L. Amendola, *Cosmic Parallax,* arXiv:astro-ph/0809.3675 (2008).

[464] S. Rasanen, *Dark energy from backreaction,* JCAP **0402** (2004), 003.

[465] 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.

[466] E. W. Kolb, S. Matarrese and A. Riotto, *On cosmic acceleration without dark energy,* New J. Phys. **8** (2006), 322.

[467] C. M. Hirata and U. Seljak, *Can superhorizon cosmological perturbations explain the acceleration of the universe?,* Phys. Rev. D **72** (2005), 083501.

[468] P. Martineau and R. H. Brandenberger, *The effects of gravitational back-reaction on cosmological perturbations,* Phys. Rev. D **72** (2005), 023507.

[469] 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.

[470] 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.

[471] T. Buchert, *Dark energy from structure - A status report,* Gen. Rel. Grav. **40** (2008), 467.

[472] S. Rasanen, *The effect of structure formation on the expansion of the universe,* Int. J. Mod. Phys. D **17** (2009), 2543.

[473] E. W. Kolb, V. Marra and S. Matarrese, *Cosmological background solutions and cosmological backreactions,* arXiv:astro-ph/0901.4566 (2009).

[474] F. Piazza, *Modifying Gravity in the Infra-Red by imposing an 'Ultra-Strong' Equivalence Principle,* arXiv:hep-th/0904.4299 (2009).

[475] F. Piazza, *The IR-Completion of Gravity: What happens at Hubble Scales?,* arXiv:hep-th/0907.0765 (2009).

[476] V. Silveira and I. Waga, *Decaying Lambda cosmologies and power spectrum,* Phys. Rev. D **50** (1994), 4890.

[477] L. M. Wang and P. J. Steinhardt, *Cluster abundance constraints on quintessence models,* Astrophys. J. **508** (1998), 483.

[478] Y. Gong, M. Ishak and A. Wang, *Growth factor parametrization in curved space,* Phys. Rev. D **80** (2009), 023002.

[479] E. V. Linder, *Cosmic growth history and expansion history,* Phys. Rev. D **72** (2005), 043529.

[480] C. Di Porto and L. Amendola, *Observational constraints on the linear fluctuation growth rate,* Phys. Rev. D **77** (2008), 083508.

[481] 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.

[482] 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.

[483] L. Amendola, *Phantom energy mediates a long-range repulsive force,* Phys. Rev. Lett. **93** (2004), 181102.

[484] S. Tsujikawa, *Reconstruction of general scalar-field dark energy models,* Phys. Rev. D **72** (2005), 083152.

[485] L. Amendola and D. Tocchini-Valentini, *Baryon bias and structure formation in an accelerating universe,* Phys. Rev. D **66** (2002), 043528.

[486] L. Amendola and C. Quercellini, *Skewness as a test of the equivalence principle,* Phys. Rev. Lett. **92** (2004), 181102.

[487] W. Hu, R. Barkana and A. Gruzinov, *Cold and fuzzy dark matter,* Phys. Rev. Lett. **85** (2000), 1158.

[488] L. Amendola and R. Barbieri, *Dark matter from an ultra-light pseudo-Goldsone-boson,* Phys. Lett. B **642** (2006), 192.

[489] Y. S. Song, H. Peiris and W. Hu, *Cosmological constraints on $f(R)$ acceleration models,* Phys. Rev. D **76** (2007), 063517.

[490] S. Tsujikawa and T. Tatekawa, *The effect of modified gravity on weak lensing,* Phys. Lett. B **665** (2008), 325.

[491] F. Schmidt, *Weak lensing probes of modified gravity,* Phys. Rev. D **78** (2008), 043002.

[492] P. Macdonald *et. al.,* *The linear theory power spectrum from the Lyman-alpha forest in the Sloan Digital Sky Survey* , Astrophys. J. **63** 5 (2005), 761.

[493] 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.

[494] K. Koyama and F. P. Silva, *Non-linear interactions in a cosmological background in the DGP braneworld,* Phys. Rev. D **75** (2007), 084040.

[495] 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.

[496] L. Amendola, M. Kunz and D. Sapone, *Measuring the dark side (with weak lensing)* , JCAP **0804** (2008), 013.

[497] S. Tsujikawa, *Matter density perturbations and effective gravitational constant in modified gravity models of dark energy,* Phys. Rev. D **76** (2007), 023514.

[498] T. Tatekawa, *Lagrangian perturbation theory in Newtonian cosmology,* Recent Res. Devel. Phys. **2** (2005), 1 [arXiv:astro-ph/0412025].

[499] F. Bernardeau, *Skewness and Kurtosis in large scale cosmic fields* , Astrophys. J. **433** (1994), 1.

[500] B. Jain and E. Bertschinger, *Second order power spectrum and non-linear evolution at high redshift* , Astrophys. J. **431** (1994), 495.

[501] J. N. Fry, *The Galaxy correlation hierarchy in perturbation theory* , Astrophys. J. **279** (1984), 499.

[502] 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.

[503] 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.

[504] F. Bernardeau, S. Colombi, E. Gaztanaga and R. Scoccimarro, *Large-scale structure of the universe and cosmological perturbation theory* , Phys. Rept. **367** (2002), 1.

[505] M. Kamionkowski and A. Buchalter, *Weakly non-linear clustering for arbitrary expansion histories,* Astrophys. J. **514** (1999), 7.

[506] E. Gaztanaga and J. A. Lobo, *Non-linear gravitational growth of large scale structures inside and outside standard cosmology,* Astrophys. J. **548** (2001), 47.

[507] R. Juszkiewicz, F. R. Bouchet and S. Colombi, *Skewness induced by gravity,* Astrophys. J. **412** (1993), L9.

[508] K. Benabed and F. Bernardeau, *Testing quintessence models with large-scale structure growth,* Phys. Rev. D **64** (2001), 083501.

[509] T. Tatekawa and S. Tsujikawa, *Second-order matter density perturbations and skewness in scalar-tensor modified gravity models,* JCAP **0809** (2008), 009.

[510] S. Matarrese and M. Pietroni, *Resumming Cosmic Perturbations,* JCAP **0706** (2007), 026.

[511] M. Crocce and R. Scoccimarro, *Renormalized cosmological perturbation theory,* Phys. Rev. D **73** (2006), 063519.

[512] T. Matsubara, *Nonlinear perturbation theory with halo bias and redshift-space distortions via the Lagrangian picture,* Phys. Rev. D **78** (2008), 083519.

[513] S. Matarrese, L. Verde and A. F. Heavens, *Large-scale bias in the universe: bispectrum method* , Mon. Not. Roy. Astron. Soc. **290** (1997), 651.

[514] 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.

[515] 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.

[516] W. H. Press and P. Schechter, *Formation of galaxies and clusters of galaxies by selfsimilar gravitational condensation,* Astrophys. J. **187** (1974), 425.

[517] J. A. Peacock, *Cosmological Physics* , Cambridge University Press, Cambridge (1999).

[518] N. N. Weinberg and M. Kamionkowski, *Constraining dark energy from the abundance of weak gravitational lenses,* Mon. Not. Roy. Astron. Soc. **341** (2003), 251.

[519] D. F. Mota and C. van de Bruck, *On the spherical collapse model in dark energy cosmologies,* Astron. Astrophys. **421** (2004), 71.

[520] I. Maor and O. Lahav, *On virialization with dark energy,* JCAP **0507** (2005), 003.

[521] 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.

[522] R. Mainini and S. Bonometto, *Mass functions in coupled dark energy models,* Phys. Rev. D **74** (2006), 043504.

[523] J. R. Bond, S. Cole, G. Efstathiou and N. Kaiser, *Excursion set mass functions for hierarchical Gaussian fluctuations,* Astrophys. J. **379** (1991), 440.

[524] R. K. Sheth and G. Tormen, *Large scale bias and the peak background split,* Mon. Not. Roy. Astron. Soc. **308** (1999), 119.

[525] S. Sasaki, *Formation rate of bound objects in the hierarchical clustering model,* Pub. Astr. Soc. Japan **46** (1994), 427.

[526] A. Jenkins *et al.* , *Mass function of dark matter halos,* Mon. Not. Roy. Astron. Soc. **321** (2001), 372.

[527] P. Rosati, S. Borgani and C. Norman, *The evolution of X-ray clusters of galaxies,* Ann. Rev. Astron. Astrophys. **40** (2002), 539.

[528] J. L. Tinker *et al.* , *Toward a halo mass function for precision cosmology: the limits of universality,* Astrophys. J. **688** (2008), 709.

[529] 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.

[530] J. A. Peacock and S. J. Dodds, *Non-linear evolution of cosmological power spectra,* Mon. Not. Roy. Astron. Soc. **280** (1996), L19.

[531] 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.

[532] A. Klypin, A. V. Maccio, R. Mainini and S. A. Bonometto, *Halo properties in models with dynamical dark energy,* Astrophys. J. **599** (2003), 31.

[533] 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.

[534] L. Casarini, A. V. Maccio and S. A. Bonometto, *Dynamical dark energy simulations: high accuracy power spectra at high redshift,* JCAP **0903** (2009), 014.

[535] 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.

[536] M. Baldi, V. Pettorino, G. Robbers and V. Springel, *N-body simulations of coupled dark energy cosmologies,* arXiv:astro-ph/0812.3901 (2008).

[537] I. Laszlo and R. Bean, *Non-linear growth in modified gravity theories of dark energy,* Phys. Rev. D **77** (2008), 024048.

[538] H. Oyaizu, *Non-linear evolution of $f(R)$ cosmologies I: methodology* , Phys. Rev. D **78** (2008), 123523.

[539] H. Oyaizu, M. Lima and W. Hu, *Non-linear evolution of $f(R)$ cosmologies II: power spectrum,* Phys. Rev. D **78** (2008), 123524.

[540] 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.

[541] B. Li and H. Zhao, *Structure formation by a fifth force: N-body versus linear simulations,* Phys. Rev. D **80** (2009), 044027.

[542] R. Kass and A. Raftery, *Bayes factors and model uncertainty* , J. Amer. Statist. Assoc. **90** (1995), 773.

[543] R. Trotta, *Bayes in the sky: Bayesian inference and model selection in cosmology,* Contemp. Phys. **49** (2008), 71.

[544] E. V. Linder and R. Miquel, *Cosmological Model Selection:. Statistics and Physics* , Int. J. Mod. Phys. D **17** (2008), 2315.

[545] 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).

[546] A. Albrecht, *et al., Report of the dark energy task force* , arXiv:astro-ph/0609591 (2006).

[547] M. Tegmark, *Measuring cosmological parameters with galaxy surveys* , Phys. Rev. Lett. **79** (1997), 3806.

[548] 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.

[549] 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.

[550] D. J. Eisenstein, W. Hu and M. Tegmark, *Cosmic Complementarity: Joint parameter estimation from CMB experiments and redshift surveys,* Astrophys. J. **518** (1999), 2.

[551] D. Huterer and G. Starkman, *Parameterization of dark-energy properties: A principal-component approach,* Phys. Rev. Lett. **90** (2003), 031301.

[552] A. J. Albrecht and G. Bernstein, *Evaluating dark energy probes using multidimensional dark energy parameters,* Phys. Rev. D **75** (2007), 103003.

[553] A. J. Albrecht *et al.* , *Findings of the joint dark energy mission figure of Merit Science Working Group,* arXiv:astro-ph/0901.0721 (2009).

[554] R. G. Crittenden and N. Turok, *Looking for Λ with the Rees-Sciama effect,* Phys. Rev. Lett. **76** (1996), 575.

[555] J. D. Jackson, *Classical Electrodynamics* , Wiley and Sons, New York, 3-rd edition (2002).

[556] 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.

[557] J. Garriga, L. Pogosian and T. Vachaspati, *Forecasting cosmic doomsday from CMB/LSS cross-correlations,* Phys. Rev. D **69** (2004), 063511.

[558] A. Cooray, *Integrated Sachs-Wolfe effect: Large scale structure correlation* , Phys. Rev. D **65** (2002), 103510.

[559] J. Garriga, L. Pogosian and T. Vachaspati, *Forecasting Cosmic Doomsday from CMB/LSS Cross-Correlations,* Phys. Rev. D **69** (2004), 063511.

[560] 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.

[561] 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.

[562] M. Douspis, P. G. Castro, C. Caprini and N. Aghanim, *Optimising large galaxy surveys for ISW detection,* arXiv:astro-ph/0802.0983 (2008).

[563] J. Weller and A. M. Lewis, *Large scale cosmic microwave background anisotropies and dark energy* , Mon. Not. Roy. Astron. Soc. **346** (2003), 987.

[564] H. J. Seo and D. J. Eisenstein, *Improved forecasts for the baryon acoustic oscillations and cosmological distance scale,* Astrophys. J. **665** (2007), 14.

[565] C. Alcock and B. Paczynski, *An evolution free test for non-zero cosmological constant,* Nature **281** (1979), 358.

[566] W. Balliger, J. A. Peacock and A. F. Heavens, *Measuring the cosmological constant with redshift surveys,* Mon. Not. Roy. Astron. Soc. **282** (1996), 877.

[567] A. Rassat *et al.* , *Deconstructing baryon acoustic oscillations: A comparison of methods,* arXiv:astro-ph/0810.0003 (2008).

[568] 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.

[569] 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.

[570] D. Dolney, B. Jain and M. Takada, *Baryon oscillations and dark-energy constraints from imaging surveys,* Mon. Not. Roy. Astron. Soc. **366** (2006), 884.

[571] 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.

[572] 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.

[573] L. Guzzo *et al.* , *A test of the nature of cosmic acceleration using galaxy redshift distortions,* Nature **451** (2008), 541.

[574] A. J. S. Hamilton, *Measuring Ω and the real correlation function from the redshift correlation function* , Astrophys. J. **385** (1992), L5.

[575] 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.

[576] M. Bartelmann and P. Schneider, *Weak Gravitational Lensing,* Phys. Rept. **340** (2001), 291.

[577] 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.

[578] 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.

[579] N. Kaiser, G. Wilson and G. A. Luppino, *Large-scale cosmic shear measurements,* arXiv:astro-ph/0003338 (2000).

[580] 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.

[581] 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.

[582] H. Hoekstra and B. Jain, *Weak Gravitational lensing and its cosmological applications* , ARNPS **58** (2008), 99.

[583] A. Amara and A. Refregier, *Optimal surveys for weak lensing tomography,* Mon. Not. Roy. Astron. Soc. **381** (2007), 1018.

[584] N. Kaiser, *Weak gravitational lensing of distant galaxies* , Astrophys. J. **388** (1992), 272.

[585] J. Miralda-Escude, *The correlation function of galaxy ellipticities produced by gravitational lensing,* Astrophys. J. **380** (1991), 1.

[586] W. Hu, *Power Spectrum Tomography with Weak Lensing* , Astrophys. J. **522** (1999), L21.

[587] C. M. Hirata and U. Seljak, *Intrinsic alignment-lensing interference as a contaminant of cosmic shear,* Phys. Rev. D **70** (2004), 063526.

[588] W. Hu and M. Tegmark, *Weak lensing: Prospects for measuring cosmological parameters,* Astrophys. J. **514** (1999), L65.

[589] M. Takada and B. Jain, *Cosmological parameters from lensing power spectrum and bispectrum tomography,* Mon. Not. Roy. Astron. Soc. **348** (2004), 897.

[590] G. M. Bernstein, *Comprehensive Two-Point Analyses of Weak Gravitational Lensing Surveys* , Astrophys. J. **695** (2009), 652.

[591] L. Fu, *et al., Very weak lensing in the CFHTLS Wide: Cosmology from cosmic shear in the linear regime,* Astron. Astrophys. **479** (2008), 9.

[592] S. Borgani, *Cosmology with clusters of galaxies* arXiv:astro-ph/0605575 (2006).

[593] S. L. Shapiro and S. A. Teukolsky, *Black holes, White Dwarfs, and Neutron Stars,* Wiley, New York (1983).

[594] A. Cavaliere and R. Fusco-Femiano, *X-rays from hot plasma in clusters of galaxies,* Astron. Astrophys. **49** (1976), 137.

[595] 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.

[596] K. Pedersen and H. Dahle, *Calibration of the mass-temperature relation for clusters of galaxies using weak gravitational lensing,* arXiv:astro-ph/0603260 (2006).

[597] S. Sasaki, *A new method to estimate cosmological parameters using the baryon fraction in clusters of galaxies,* Pub. Astr. Soc. Japan **48** (1996), L119.

[598] U.-L. Pen, *Measuring the universal deceleration using angular diameter distances to clusters of galaxies* , New Astronomy, **2** (1997), 309.

[599] 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.

[600] 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.

[601] S. Ettori *et al.* , *The cluster gas mass fraction as a cosmological probe: a revised study,* arXiv:astro-ph/0904.2740 (2009).

[602] 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).

[603] 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).

[604] G. Ghirlanda, G. Ghisellini and C. Firmani, *Gamma Ray Bursts as standard candles to constrain the cosmological parameters,* New J. Phys. **8** (2006), 123.

[605] 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.

[606] A. P. Bouvier, V. Petrosian and F. Ryde, *On the use of gamma-ray bursts as a cosmological tool,* American Astronomical Society, HEAD meeting, Bulletin of the American Astronomical Society, **38** (2006), 369.

[607] S. Basilakos and L. Perivolaropoulos, *Testing GRBs as standard candles,* Mon. Not. Roy. Astron. Soc. **391** (2008), 411.

[608] 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.

[609] 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.

[610] 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.

[611] J. Simon, L. Verde and R. Jimenez, *Constraints on the redshift dependence of the dark energy potential,* Phys. Rev. D **71** (2005), 123001.

[612] M. A. Dantas and J. S. Alcaniz, *Current lookback time-redshift bounds on dark energy,* arXiv:astro-ph/0901.2327 (2009).

[613] P. Schneider, J. Ehlers and E. E. Falco, *Gravitational Lenses* , Springer-Verlag, Berlin (1992).

[614] 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.

[615] C. Porciani, and P. Madau, *Gravitational Lensing of Distant Supernovae in Cold Dark Matter Universes,* Astrophys. J. **532** (2000), 679.

[616] L. X. Li and J. P. Ostriker, *Semi-analytical models for lensing by dark halos: I. Splitting angles* , Astrophys. J. **566** (2002), 652.

[617] J. F. Navarro, C. S. Frenk and S. D. M. White, *The structure of cold dark matter Halos,* Astrophys. J. **462** (1996), 563.

[618] H. Zhao, *Analytical models for galactic nuclei,* Mon. Not. Roy. Astron. Soc. **278** (1996), 488.

[619] 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.

[620] A. S. Bolton *et al.* , *The Sloan Lens ACS Survey. V. The full ACS strong-lens sample,* Astrophys. J. **682** (2008), 964.

[621] 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.

[622] M. Bartelmann, M. Meneghetti, F. Perrotta, C. Baccigalupi and L. Moscardini, *Arc statistics in cosmological models with dark energy,* Astron. Astrophys. **409** (2003), 449.

[623] 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.

[624] 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.

[625] 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.

[626] 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.

[627] A. Sandage, *The change of redshift and apparent luminosity of galaxies due to the deceleration of selected expanding universes* , Astrophys. J. **136** (1962), 319.

[628] 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.

[629] K. Lake, *Comment on the time evolution of the cosmological redshift* , Astrophys. J. **247** (1981), 17.

[630] R. Rudiger, *Time variation of the cosmological redshift in Dicke-Brans-Jordan cosmologies* , Astrophys. J. **260** (1982), 33.

[631] A. Loeb, *Direct measurement of cosmological parameters from the cosmic deceleration of extragalactic objects* arXiv:astro-ph/9802122 (1998).

[632] J. Liske *et al.* , *Cosmic dynamics in the era of extremely large telescopes* , Mon. Not. Roy. Astron. Soc. Lett. **386** (2008), 1192.

[633] 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.

[634] 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.

[635] 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.

[636] T. Koivisto and D. F. Mota, *Anisotropic dark energy: dynamics of the background and perturbations* , JCAP **6** (2008), 18.

[637] D. C. Rodrigues, *Anisotropic cosmological constant and the CMB quadrupole anomaly* , Phys. Rev. D **77** (2008), 023534.

[638] Gaia website at http://www.rssd.esa.int/Gaia/

[639] 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).

[640] B. A. Bassett and M. Kunz, *Is cosmic distance-duality violated?* Phys. Rev. D **69** (2004), 101305.

[641] A. Avgoustidis, L. Verde and R. Jimenez, *Consistency among distance measurements: transparency, BAO scale and accelerated expansion,* arXiv:astro-ph/0902.2006 (2009).

[642] 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.

[643] P. J. E. Peebles, *Tests of Cosmological Models Constrained by Inflation,* Astrophys. J. **284** (1984), 439.

[644] 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.

[645] N. Sugiyama, N. Gouda, and M. Sasaki, *Constraints on universe models with cosmological constant from cosmic microwave background anisotropy* , Astrophys. J. **365** (1990), 432.

[646] L M. Krauss and M. S. Turner, *The cosmological constant is back,* Gen. Rel. Grav. **27** (1995), 1137.

[647] J. Frieman, M. Turner and D. Huterer, *Dark energy and the accelerating universe,* Ann. Rev. Astron. Astrophys. **46** (2008), 385.

[648] M. Spinelli, A. Zoldan, G. De Lucia, L. Xie, and M. Viel, *The atomic hydrogen content of the post-reionization Universe,* Mon. Not. Roy. Astron. Soc. **493** 4 (2020), 5434

[649] A. Cooray, R. Sheth, *Halo models of large scale structure,* Phys. Rep. **372** 1 (2002), 1

[650] P. Schechter, *An analytic expression for the luminosity function for galaxies,* Astroph. J. **203** (1976), 297

[651] E. E. O. Ishida, S. D. P. Vitenti and M. Penna-Lima and J. Cisewski and R. S. de Souza and A. M. M. Trindade and E. Cameron and V. C. Busti and COIN Collaboration, *COSMOABC: Likelihood-free inference via Population Monte Carlo Approximate Bayesian Computation,* Astron. and Comp. **13** (2015).

[652] N. Metropolis, A. W. Rosenbluth and M. N. Rosenbluth and A. H. Teller and E. Teller, *Equation of State Calculations by Fast Computing Machines,* J. of Chem. Phys. **21** (1953).

[653] K. Fukushima, *Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position,* Biolog. Cybernetics. **36** (1980), 193.

[654] N. Jeffrey and B. D. Wandelt, *Solving high-dimensional parameter inference: marginal posterior densities & Moment Networks,* Third Workshop on Machine Learning and the Physical Sciences, NeurIPS (2020)

[655] S. Kullback and R. A. Leibler, *On Information and Sufficiency,* The Annals of Mathematical Statistic **22,** 1 (1951)

[656] L. K. Saul, T. Jaakkola and M. T. Jordan, *Mean Field Theory for Sigmoid Belief Networks,* Journal of Artificial Intelligence Research **4** (1996)

[657] C. Blundell, J. Cornebise, K. Kavukcuoglu and D. Wierstra, *Weight Uncertainty in Neural Networks,* Proceedings of the 32nd International Conference on Machine Learning, ICML 2015 (2015)

[658] D. P. Kingma, T. Salimans and M. Welling, *Variational Dropout and the Local Reparameterization Trick,* Advances in neural information processing systems **28** (2015)

[659] J. Alsing, T. Charnock, S. Feeney and B. Wandelt, *Fast likelihood-free cosmology with neural density estimators and active learning,* Mon. Not. Roy. Astron. Soc. **488** 3 (2019)

[660] L. Ardizzone, J. Kruse, S. Wirkert, D. Rahner, E. W. Pellegrini, R. S. Klessen, L. Maier-Hein, C. Rother, and U. Köthe, *Analyzing Inverse Problems with Invertible Neural Networks,* conference paper at ICLR (2019)

[661] S. T. Radev, U. K. Mertens, A. Mertens, L. Ardizzone, and U. Köthe, *BayesFlow: Learning complex stochastic models with invertible neural networks,* IEEE Transactions on Neural Networks and Learning Systems **33** 4 (2022)