

1

A short review of standard and inflationary cosmology

In this chapter we will recall some basic notions of standard and inflationary cosmology that will be used later, in a string cosmology context. We will assume that the reader is already familiar with the geometric formalism of the theory of general relativity, and with the main observational aspects of large-scale astronomy and astrophysics. We will discuss, in particular, the various assumptions of the so-called standard cosmological model, the problems associated with its initial conditions, and the basic aspects of its “inflationary” completion driven by the potential energy of a cosmic scalar field (further details on the inflationary scenario will be supplied in Chapter 8). This presentation aims at a self-contained study of the early cosmological dynamics: for a more detailed introduction, and a deeper analysis of the topics discussed in this chapter, we refer the interested reader to [1, 2, 3] for the standard cosmological model, and to [4, 5, 6] for the inflationary scenario.

1.1 The standard cosmological model

The standard cosmological model, developed during the second half of the last century, was inspired by two fundamental observational results: the recession of galaxies, discovered by Hubble [7], and the presence of the Cosmic Microwave Background (CMB), discovered by Penzias and Wilson [8]. The model relies upon a number of hypotheses – also motivated by direct and indirect observations – that we now list, with some illustrative discussion.

1.1.1 Einstein equations

The first assumption is that the gravitational interaction, on cosmological scales of distance, is well described by the classical theory of general relativity,

and in particular by the equations derived from the effective four-dimensional action

$$S = -\frac{1}{16\pi G} \int d^4x \sqrt{-g} R + S_\Sigma + \int d^4x \sqrt{-g} \mathcal{L}_m. \quad (1.1)$$

Here S_Σ is the Gibbons–Hawking boundary term [9], required in order to reproduce the standard Einstein equations, and $\mathcal{L}_m$ is the Lagrangian density of the matter fields, acting as gravitational sources. The variation of the action (1.1) with respect to the metric $g_{\mu\nu}$ yields (see Chapter 2 for an explicit derivation)

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi G T_{\mu\nu}, \quad (1.2)$$

where $G_{\mu\nu}$ is the so-called Einstein tensor, and $T_{\mu\nu}$ is the (dynamical) energy-momentum tensor of the matter sources, defined by the variation (or functional differentiation) of the matter action as

$$\delta_g (\sqrt{-g} \mathcal{L}_m) = \frac{1}{2} \sqrt{-g} T_{\mu\nu} \delta g^{\mu\nu}. \quad (1.3)$$

The right-hand side of Eq. (1.2) represents all the sources gravitationally coupled to the metric, and therefore includes the possible contribution of the vacuum energy density associated with a cosmological constant Λ , and described by the effective energy-momentum tensor $T_{\mu\nu} = \Lambda g_{\mu\nu}$.

1.1.2 Homogeneity and isotropy

A second assumption is that the spatial sections of the Universe, on large enough scales of distance, can be described as homogeneous and isotropic (three-dimensional) Riemann manifolds, geometrically represented by maximally symmetric spaces where rotations and translations form a six-parameter isometry group.

It may be noted that, on scales much smaller than the Hubble radius $H_0^{-1} \simeq 0.9h^{-1} \times 10^{28}$ cm, the distribution of visible matter seems to follow a “fractal” distribution (see for instance [10]), and that it is not very clear, at present, at which scale the (averaged) matter distribution becomes really homogeneous and isotropic. The hypothesis of homogeneity and isotropy refers, however, to the full set of cosmic gravitational sources (including, as we shall see, radiation, dark matter, dark energy, ...), and is quite powerful, since it allows a simplified cosmological description in which the space-time geometry can be parametrized by the so-called “comoving” chart (or set of coordinates). In that case, the fundamental space-time interval reduces to

$$ds^2 = b^2(t) dt^2 - a^2(t) d\sigma^2(\vec{r}), \quad (1.4)$$

1.1 The standard cosmological model

3

where $a(t)$, $b(t)$ are generic functions of the time coordinate, and $d\sigma^2$ is the line-element of a three-dimensional space with constant (positive, negative or zero) curvature K . Using a set of stereographic coordinates $\{x_1, x_2, x_3\}$, the metric of such a maximally symmetric space can be parametrized as [1]

$$d\sigma^2 = dx_i dx^i + K \frac{(x_i dx^i)^2}{1 - K x_i x^i}, \quad (1.5)$$

where scalar products are performed with the Euclidean metric δ_{ij} .

An important property of the comoving chart is the fact that *static* observers, with four-velocity $u^\mu = (u^0, \vec{0})$, are also *geodesic* observers. The normalization condition $g_{\mu\nu} u^\mu u^\nu = 1$, with the metric (1.4), gives indeed $u^0 = b^{-1}(t)$ and

$$\frac{du^0}{d\tau} = -\frac{\dot{b}}{b^3}, \quad \Gamma_{00}^0 (u^0)^2 = \frac{\dot{b}}{b^3}, \quad (1.6)$$

which implies that the field u^0 satisfies the geodesic equation

$$\frac{du^0}{d\tau} + \Gamma_{00}^0 (u^0)^2 = 0. \quad (1.7)$$

Here τ is the proper time (related to the coordinate time t by $d\tau = \sqrt{g_{00}} dt = b(t) dt$), and the dot denotes differentiation with respect to t . In addition, if $u^i = 0$, then

$$\frac{du^i}{d\tau} = -\Gamma_{00}^i (u^0)^2 = -\frac{1}{2b^2} g^{ij} (2\partial_0 g_{j0} - \partial_j g_{00}) \equiv 0. \quad (1.8)$$

Thus, in the absence of non-gravitational forces, static observers are always at rest with respect to comoving coordinates, even if the geometry is time dependent.

The existence of such observers provides a natural reference frame for synchronizing clocks, and suggests the use of a convenient time coordinate, the so-called *cosmic time*, which corresponds to the proper time of the static observers. The choice of this time coordinate leads to the *synchronous gauge*, defined by the condition $g_{00} = 1$. It is also convenient to parametrize the maximally symmetric space of Eq. (1.5) with spherical coordinates $\{r, \theta, \varphi\}$. By setting $x_1 = r \sin \theta \cos \varphi$, $x_2 = r \sin \theta \sin \varphi$, $x_3 = r \cos \theta$, and differentiating to compute $d\sigma^2$, in the synchronous gauge of the comoving chart, one finally arrives at the well-known Robertson–Walker metric, defined by

$$ds^2 = dt^2 - a^2(t) \left[\frac{dr^2}{1 - Kr^2} + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2) \right]. \quad (1.9)$$

Here t is the cosmic-time coordinate, and the constant K (with dimensions L^{-2}) controls the intrinsic curvature of the space-like $t = \text{const}$ hypersurfaces, representing three-dimensional sections of the space-time manifold. With our conventions

the function $a(t)$, called the “scale factor”, is dimensionless, while the comoving radial coordinate r has conventional dimensions of length.

Another choice of time coordinate (often used in this book) is the so-called *conformal gauge*, defined by the condition $g_{00} = a^2$. The time parameter of this gauge, usually denoted by η , is thus related to the cosmic time t by $dt = a d\eta$. The choice of the conformal gauge is particularly convenient for spatially flat manifolds ($K = 0$), whose metric can then be written in conformally flat form, using cartesian coordinates, as

$$ds^2 = a^2(\eta) (d\eta^2 - dx_i dx^i). \quad (1.10)$$

A space-time described by the Robertson–Walker metric is characterized by a number of interesting kinematical properties concerning the motion of test bodies and the propagation of signals (see for instance [1]). For the purposes of this book it will be enough to recall two effects.

The first effect concerns the spectral shift of a periodic signal, a shift originating from the well-known temporal slow-down produced by gravity. Indeed, at any given time t , all points of the three-dimensional spatial sections at constant curvature will be affected by exactly the same gravitational field, so that any local process will be equally slowed-down with respect to the same process occurring in the flat Minkowski space, quite independently of its spatial position. However, if the scale factor $a(t)$ varies with time, then the curvature radius of the spatial sections (and the associated intensity of the local effective gravitational field) will also vary with time. This will produce a difference in the local gravitational field (and in the local “slow-down”) between the time t_{em} of emission of a periodic signal of pulsation ω_{em} , and the time $t_{\text{obs}} > t_{\text{em}}$ when the same signal is observed with pulsation ω_{obs} . The ratio of the two pulsations will be clearly proportional to the ratio of the local gravitational intensities at t_{em} and t_{obs} , and thus inversely proportional to the spatial curvature radius.

For a more precise computation of the spectral shift $\omega_{\text{em}}/\omega_{\text{obs}}$ we may consider a photon of four-momentum p^μ , traveling along a null geodesic of a spatially flat Robertson–Walker metric. In the cosmic-time gauge such a null path has differential equation $dt = a \hat{n}_i dx^i$, where $\hat{n}$ is a unit vector ($|\hat{n}| = 1$) specifying the photon direction; the null photon momentum is, in this gauge, $p^\mu = p^0(1, \hat{n}^i/a)$, with $g_{\mu\nu} p^\mu p^\nu = 0$. The momentum is parallelly transported along the geodesic, and for the energy p^0 we have, in particular,

$$\begin{aligned} dp^0 &= -\Gamma_{\alpha\beta}^0 dx^\alpha p^\beta = \Gamma_{ij}^0 dx^i p^j \\ &= -\dot{a} p^0 \hat{n}_i dx^i = -\frac{\dot{a}}{a} p^0 dt. \end{aligned} \quad (1.11)$$

The integration gives $p^0 = \bar{\omega}/a(t)$, where the integration constant $\bar{\omega}$ represents the proper frequency of the photon in the Minkowski space locally tangent to the given cosmological manifold.

The local frequency measured by a static, comoving observer u^μ is thus time dependent, being determined by the projection $p^\mu u_\mu = \bar{\omega}/a(t)$. A photon emitted at $t = t_{\text{em}}$ and received at $t = t_{\text{obs}}$, even in the absence of a (possible) Doppler effect due to the relative motion of source and emitter, will be characterized by the spectral shift

$$\frac{\omega_{\text{em}}}{\omega_{\text{obs}}} = \frac{(p^\mu u_\mu)_{\text{em}}}{(p^\mu u_\mu)_{\text{obs}}} = \frac{a_{\text{obs}}}{a_{\text{em}}} \quad (1.12)$$

(see also Eqs. (8.172)–(8.173), and the discussion of Section 8.2). If the Universe is expanding, then $a_{\text{obs}} > a_{\text{em}}$ for $t_{\text{obs}} > t_{\text{em}}$, and the Robertson–Walker metric produces an effective redshift of the signals received from distant sources, i.e. $\omega_{\text{obs}} < \omega_{\text{em}}$. In particular, since observations are carried out at the present time, $t_{\text{obs}} = t_0$, it may be useful to introduce a redshift parameter $z(t)$ defined as

$$1 + z(t) = \frac{a(t_0)}{a(t)} \equiv \frac{a_0}{a(t)}, \quad (1.13)$$

which controls the relative “stretching” of the wavelengths of the received radiation,

$$z = \frac{\Delta\lambda}{\lambda} = \frac{\lambda_{\text{obs}} - \lambda_{\text{em}}}{\lambda_{\text{em}}}. \quad (1.14)$$

A second important feature of the Robertson–Walker kinematics, which we recall here for later applications, is the possible existence of “horizons”, i.e. of surfaces with relevant causal properties. For any given observer we may consider, in particular, the *particle horizon*, which divides the portion of space-time already observed from the one yet to be observed, and the *event horizon*, which divides the observable portion of space-time from the one causally disconnected [11]. For their precise definition we must refer to the limiting times t_{m} and t_{M} corresponding, respectively, to the maximum *past* extension and *future* extension of the time coordinate on the given cosmological manifold.

Let us consider a signal propagating towards the origin along a null radial geodesic of the metric (1.9) ($ds^2 = 0$, $d\theta = 0 = d\varphi$), satisfying the equation $dt/a = dr/\sqrt{1 - Kr^2}$, and received by a comoving observer at rest at the origin of the polar coordinate system. A signal emitted from a radial position $r = r_1$, at a time $t = t_1$, will be received at $r = 0$ at a time $t = t_0 > t_1$, such that

$$\int_0^{r_1} \frac{dr}{\sqrt{1 - Kr^2}} = \int_{t_1}^{t_0} \frac{dt}{a(t)}. \quad (1.15)$$

The considered signal was emitted at a *proper* distance $d(t)$ from the origin which, at time t_0 , is determined by

$$d(t_0) = a(t_0) \int_0^{r_1} \frac{dr}{\sqrt{1 - Kr^2}} = a(t_0) \int_{t_1}^{t_0} \frac{dt}{a(t)}. \quad (1.16)$$

In the limit $t_1 \rightarrow t_m$ we then define the “**particle horizon**”, for the given observer at time t_0 , as the spherical surface centered at the origin $r = 0$ with proper radius

$$d_p(t_0) = a(t_0) \int_{t_m}^{t_0} \frac{dt}{a(t)}. \quad (1.17)$$

This surface encloses the maximal portion of space physically accessible to direct observation from the origin of the coordinate system at the time t_0 . Points located at a proper spatial distance $d > d_p(t_0)$ *cannot* be causally connected with the given observer at the given time t_0 (they may become causally connected at later times, at least in principle).

Consider now a radial signal emitted towards the origin at time t_0 , from a point located at a comoving position r_2 , and received at the origin at a time $t_2 > t_0$. The proper distance of the emitter from the origin, at time t_0 , is then

$$d(t_0) = a(t_0) \int_0^{r_2} \frac{dr}{\sqrt{1 - Kr^2}} = a(t_0) \int_{t_0}^{t_2} \frac{dt}{a(t)}. \quad (1.18)$$

In the limit $t_2 \rightarrow t_M$ we can then define the “**event horizon**”, at the time t_0 , as the spherical surface centered at the origin with proper radius

$$d_e(t_0) = a(t_0) \int_{t_0}^{t_M} \frac{dt}{a(t)}. \quad (1.19)$$

Signals emitted from points located at a proper distance $d > d_e(t_0)$ will *never* be able to reach the origin. In other words, points with spatial separations $d > d_e$ will never become causally connected, even extending the time coordinate to the extremal future limit allowed by the given cosmological manifold.

The above horizons exist if the integrals of Eqs. (1.17) and (1.19) are convergent, of course. Consider, for instance, a cosmological solution describing a Universe expanding for ever from an initial singularity, and parametrized in cosmic time by the power-law scale factor $a(t) = t^\alpha$, with $\alpha > 0$, and $0 \leq t \leq \infty$: it can be easily checked that the particle horizon exists if $0 < \alpha < 1$, while the event horizon exists if $\alpha > 1$. For $\alpha = 1$ neither the particle horizon nor the event horizon exists. The definitions of horizon given here will be used in the following chapters, and will be applied in particular in Section 5.3 to illustrate some important differences between standard and string cosmology models of inflation.

1.1.3 Perfect fluid sources

A third assumption (or set of assumptions) of the standard cosmological model refers to the gravitational sources that we need to specify in order to solve the Einstein equations. According to the standard model the sources of the cosmological gravitational field on large scales, after averaging over possible spatial fluctuations, can be represented as a barotropic, perfect fluid with energy-momentum tensor

$$T_{\mu}^{\nu} = (\rho + p)u_{\mu}u^{\nu} - p\delta_{\mu}^{\nu}, \quad (1.20)$$

where the energy density ρ and pressure p depend only on time, and are related by the equation of state

$$\frac{p}{\rho} = \gamma = \text{const.} \quad (1.21)$$

In addition, the fluid is assumed to be at rest in the comoving frame. Thus, in the synchronous gauge, $u^{\mu} = (1, \vec{0})$ and T_{μ}^{ν} becomes diagonal,

$$T_0^0 = \rho(t), \quad T_i^j = -p(t)\delta_i^j. \quad (1.22)$$

With the given sources we are now able to write explicitly the Einstein equations (1.2), using the following (more convenient, but equivalent) form:

$$R_{\mu}^{\nu} = 8\pi G \left(T_{\mu}^{\nu} - \frac{1}{2}T\delta_{\mu}^{\nu} \right). \quad (1.23)$$

For the Robertson–Walker metric (1.9) the non-zero components of the Ricci tensor, in mixed form, depend only on time, and are given by

$$\begin{aligned} R_1^1 = R_2^2 = R_3^3 &= -\frac{\ddot{a}}{a} - 2 \left(H^2 + \frac{K}{a^2} \right), \\ R_0^0 &= -3\frac{\ddot{a}}{a}, \end{aligned} \quad (1.24)$$

where $H = \dot{a}/a$ (the dot indicates the derivative with respect to cosmic time). The time and spatial components of Eqs. (1.23) then provide, respectively, the following independent equations:

$$\begin{aligned} \frac{\ddot{a}}{a} &= -\frac{4\pi G}{3}(\rho + 3p), \\ \frac{\ddot{a}}{a} + 2 \left(H^2 + \frac{K}{a^2} \right) &= 4\pi G(\rho - p). \end{aligned} \quad (1.25)$$

Combining them in order to eliminate $\ddot{a}/a$, and differentiating the energy density ρ with respect to time, leads to the system of first-order differential equations:

$$H^2 + \frac{K}{a^2} = \frac{8\pi G}{3}\rho, \quad (1.26)$$

$$\dot{\rho} + 3H(\rho + p) = 0. \quad (1.27)$$

The last equation can also be directly obtained from the covariant conservation of the energy-momentum tensor, $\nabla_\nu T^\nu_\mu = 0$, which is a consequence of the contracted Bianchi identity $\nabla_\nu G^\nu_\mu = 0$ (see Eq. (1.2)).

In order to solve the above system of equations for the three unknown functions $a(t)$, $\rho(t)$, $p(t)$, it is necessary to use the equation of state $p = p(\rho)$, which in our case corresponds to the barotropic condition (1.21). In general, the gravitational sources of the standard cosmological model can be represented as a mixture of barotropic perfect fluids,

$$\rho = \sum_n \rho_n, \quad p = \sum_n p_n, \quad p_n = \gamma_n \rho_n, \quad (1.28)$$

with no energy transfer between the different fluid components, so that the energy-momentum tensor of each fluid is separately conserved. Equation (1.27) then yields, for each component,

$$\rho_n(t) = \rho_n(t_0) \left(\frac{a}{a_0} \right)^{-3(1+\gamma_n)}, \quad (1.29)$$

where $\rho_n(t_0)$ is an integration constant. Since the energy density of the different components has a different time behavior, the evolution of the Universe will then be characterized by different phases, each of them dominated by different fluid components.

In each cosmological phase the time evolution of the scale factor can be obtained by substituting Eq. (1.29) into (1.26), and solving the corresponding differential equation for $a(t)$. If, in particular, we are interested in the very early time evolution we can neglect the spatial curvature term (see below), and we obtain the scale factor

$$a_n(t) = \left(\frac{t}{t_0} \right)^{2/3(1+\gamma_n)}, \quad \gamma_n \neq -1, \quad (1.30)$$

where t_0 is an integration constant. The case $\gamma_n = -1$ corresponds to the energy-momentum tensor of a cosmological constant

$$T^\nu_\mu = \Lambda \delta^\nu_\mu, \quad (1.31)$$

1.1 The standard cosmological model

9

which describes an effective fluid with equation of state $p_n = -\rho_n = -\Lambda = \text{const}$ (see Eq. (1.22)). In this case Eq. (1.29) is still valid, and the integration of Eq. (1.26) (with $K = 0$) gives the exponential solution

$$a_n(t) = \exp[H(t - t_0)], \quad H = \left(\frac{8\pi G \Lambda}{3} \right)^{1/2} = \text{const.} \quad (1.32)$$

The standard cosmological model, in its original formulation [1], assumes that the cosmic fluid consists of two fundamental components: incoherent matter (ρ_m) with zero pressure $p_m = 0$, and radiation (ρ_r) with pressure $p_r = \rho_r/3$. The radiation component of the cosmic fluid represents the contribution of all massless (or very light) relativistic particles (photons, gravitons, neutrinos, ...), while the pressureless matter component takes into account the large-scale contribution of the macroscopic gravitational sources (galaxies, clusters, interstellar gas, ...), and the contribution of cosmic backgrounds of heavy, non-relativistic particles (baryons, as well as other, more exotic, possible dark-matter components). As we shall see later in more detail (see Eq. (1.39)), the present energy density of incoherent matter is roughly of the same order of magnitude as the critical density, $\rho_m(t_0) \sim \rho_c(t_0)$, where [12]

$$\rho_c(t_0) = \frac{3 H_0^2}{8\pi G} = 3 H_0^2 M_P^2 \simeq 2.25 h^2 \times 10^{-120} M_P^4, \quad (1.33)$$

and is thus much greater than the radiation energy density today, since [12]

$$\rho_r(t_0) \simeq 4.15 h^{-2} \times 10^{-5} \rho_c(t_0). \quad (1.34)$$

Therefore, according to the standard cosmological model, the present scale factor (assuming negligible spatial curvature) should evolve in time as $a(t) \sim t^{2/3}$.

As the Universe expands, however, the energy density of the matter component decreases in time as the inverse of the proper volume, $\rho_m \sim a^{-3}$, i.e. more slowly than the radiation component, $\rho_r \sim a^{-4}$ (see Eq. (1.29)). Going backwards in time one thus necessarily reaches the so-called *equality* time, $t = t_{\text{eq}}$, characterized by the same amount of matter and radiation energy density, $\rho_m(t_{\text{eq}}) = \rho_r(t_{\text{eq}})$. At earlier times, $t < t_{\text{eq}}$, the standard model then predicts the existence of a primordial phase where the radiation is the dominant component of the total energy density, and the scale factor evolves with different kinematics, $a(t) \sim t^{1/2}$, according to Eq. (1.30).

It is worth stressing that both the matter-dominated and the radiation-dominated regimes, according to the standard model, correspond to a phase of expansion which is *decelerated* and has *decreasing curvature*, i.e. satisfies

$$\dot{a} > 0, \quad \ddot{a} < 0, \quad \dot{H} < 0, \quad (1.35)$$

as one can easily verify by differentiating Eq. (1.30) for $\gamma = 0$ and $\gamma = 1/3$ (with a power-law scale factor, we can take H as a good indicator of the time behavior of the space-time curvature scale). However, the recent large-scale observations concerning both the Hubble diagram of Type Ia Supernovae [13, 14] and the harmonic analysis of the CMB anisotropies [15, 16, 17] seem to indicate, with a growing level of precision and confidence [18, 19, 20], that the present Universe is undergoing a phase of accelerated expansion, $\ddot{a} > 0$.

Such observations are thus compatible with the first of Eqs. (1.25) only if the sources of cosmic gravity are presently dominated by a component with negative enough pressure (i.e. $\rho + 3p < 0$), so as to produce a kind of “cosmic repulsion” on large scales. Adding explicitly this new source ρ_q (dubbed “*quintessence*”, or “*dark energy*”) to the usual dust matter sources ρ_m , Eq. (1.26) becomes

$$H^2 + \frac{K}{a^2} = \frac{8\pi G}{3}(\rho_m + \rho_q), \quad (1.36)$$

where $\rho_q > \rho_m$, and $p_q/\rho_q \equiv \gamma_q < -1/3$. Dividing by H^2 we can then obtain a relation between the various components of the cosmic fluid in critical units, i.e.

$$1 = \Omega_m + \Omega_q + \Omega_K, \quad (1.37)$$

where

$$\Omega_m = \frac{\rho_m}{\rho_c}, \quad \Omega_q = \frac{\rho_q}{\rho_c}, \quad \Omega_K = -\frac{K}{a^2 H^2}. \quad (1.38)$$

The simplest model of dark energy is a cosmological constant, $\rho_q = \Lambda = \text{const}$ (which corresponds to $\gamma_q = -1$). In this case, replacing Ω_q with $\Omega_\Lambda = \Lambda/\rho_c$, the results of present observations can be summarized as follows [12]:

$$\Omega_m = 0.24^{+0.03}_{-0.04}, \quad \Omega_\Lambda = 0.76^{+0.04}_{-0.06}. \quad (1.39)$$

These results refer to the particular case $K = 0$, but can be consistently applied to the present cosmological state where the allowed deviations of $\Omega_m + \Omega_\Lambda$ from 1 are very small: indeed,

$$\Omega_K = -0.015^{+0.020}_{-0.016} \quad (1.40)$$

according to a recent combination of supernovae and CMB data [20].

The experimental results are not very different from those of Eq. (1.39) even if ρ_q does not correspond to a cosmological constant, but represents the contribution of some weakly coupled, time-dependent field, as will be discussed in Section 9.3. In such a case, the effective equation of state $\gamma_q = p_q/\rho_q$ of the dark-energy component is presently constrained by the limits

$$\gamma_q = -0.97^{+0.07}_{-0.09}, \quad (1.41)$$