Part I

Homogeneous isotropic universe

1

Kinematics and dynamics of an expanding universe

The most important feature of our universe is its large scale homogeneity and isotropy. This feature ensures that observations made from our single vantage point are representative of the universe as a whole and can therefore be legitimately used to test cosmological models.

For most of the twentieth century, the homogeneity and isotropy of the universe had to be taken as an assumption, known as the "Cosmological Principle." Physicists often use the word "principle" to designate what are at the time wild, intuitive guesses in contrast to "laws," which refer to experimentally established facts.

The Cosmological Principle remained an intelligent guess until firm empirical data, confirming large scale homogeneity and isotropy, were finally obtained at the end of the twentieth century. The nature of the homogeneity is certainly curious. The observable patch of the universe is of order 3000 Mpc (1 Mpc $\simeq 3.26 \times 10^6$ light years $\simeq 3.08 \times 10^{24}$ cm). Redshift surveys suggest that the universe is homogeneous and isotropic only when coarse grained on 100 Mpc scales; on smaller scales there exist large inhomogeneities, such as galaxies, clusters and superclusters. Hence, the Cosmological Principle is only valid within a limited range of scales, spanning a few orders of magnitude.

Moreover, theory suggests that this may not be the end of the story. According to inflationary theory, the universe continues to be homogeneous and isotropic over distances larger than 3000 Mpc, but it becomes highly inhomogeneous when viewed on scales *much much* larger than the observable patch. This dampens, to some degree, our hope of comprehending the entire universe. We would like to answer such questions as: What portion of the entire universe is like the part we find ourselves in? What fraction has a predominance of matter over antimatter? Or is spatially flat? Or is accelerating or decelerating? These questions are not only difficult to answer, but they are also hard to pose in a mathematically precise way. And, even if a suitable mathematical definition can be found, it is difficult to imagine how we could verify empirically any theoretical predictions concerning scales greatly exceeding the observable universe. The subject is too seductive to avoid speculations altogether, but we will, nevertheless, try to focus on the salient, empirically testable features of the observable universe.

It is firmly established by observations that our universe:

- *is homogeneous and isotropic on scales larger than* 100 Mpc *and has well developed inhomogeneous structure on smaller scales;*
- *expands according to the Hubble law.*

Concerning the matter composition of the universe, we know that:

- *it is pervaded by thermal microwave background radiation with temperature* $T \simeq 2.73\,\mathrm{K}$;
- *there is baryonic matter, roughly one baryon per* 10^9 *photons, but no substantial amount of antimatter;*
- *the chemical composition of baryonic matter is about* 75% *hydrogen,* 25% *helium, plus trace amounts of heavier elements;*
- *baryons contribute only a small percentage of the total energy density; the rest is a dark component, which appears to be composed of cold dark matter with negligible pressure* ($\sim$25%) *and dark energy with negative pressure* ($\sim$70%).

Observations of the fluctuations in the cosmic microwave background radiation suggest that:

- *there were only small fluctuations of order* 10^{-5} *in the energy density distribution when the universe was a thousand times smaller than now.*

For a review of the observational evidence the reader is encouraged to refer to recent papers and reviews. In this book we concentrate mostly on theoretical understanding of these basic observational facts.

Any cosmological model worthy of consideration must be consistent with established facts. While the standard big bang model accommodates most known facts, a physical theory is also judged by its predictive power. At present, inflationary theory, naturally incorporating the success of the standard big bang, has no competitor in this regard. Therefore, we will build upon the standard big bang model, which will be our starting point, until we reach contemporary ideas of inflation.

1.1 Hubble law

In a nutshell, the standard big bang model proposes that the universe emerged about 15 billion years ago with a homogeneous and isotropic distribution of matter at very high temperature and density, and has been expanding and cooling since then. We begin our account with the Newtonian theory of gravity, which captures many of the essential aspects of the universe's dynamics and gives us an intuitive grasp of what happens. After we have reached the limits of validity of Newtonian theory, we turn to a proper relativistic treatment.

In an expanding, homogeneous and isotropic universe, the relative velocities of observers obey the *Hubble law*: the velocity of observer B with respect to A is

$$\mathbf{v}_{B(A)} = H(t)\mathbf{r}_{BA}, \tag{1.1}$$

where the Hubble parameter $H(t)$ depends only on the time t, and $\mathbf{r}_{BA}$ is the vector pointing from A to B. Some refer to H as the Hubble "constant" to stress its independence of the spatial coordinates, but it is important to recognize that H is, in general, time-varying.

In a homogeneous, isotropic universe there are no *privileged* vantage points and the expansion appears the same to all observers wherever they are located. The Hubble law is in complete agreement with this. Let us consider how two observers A and B view a third observer C (Figure 1.1). The Hubble law specifies the velocities of the other two observers relative to A:

$$\mathbf{v}_{B(A)} = H(t)\mathbf{r}_{BA}, \quad \mathbf{v}_{C(A)} = H(t)\mathbf{r}_{CA}. \tag{1.2}$$

From these relations, we can find the relative velocity of observer C with respect to observer B:

$$\mathbf{v}_{C(B)} = \mathbf{v}_{C(A)} - \mathbf{v}_{B(A)} = H(t)(\mathbf{r}_{CA} - \mathbf{r}_{BA}) = H\mathbf{r}_{CB}. \tag{1.3}$$

The result is that observer B sees precisely the same expansion law as observer A. In fact, the Hubble law is the *unique* expansion law compatible with homogeneity and isotropy.

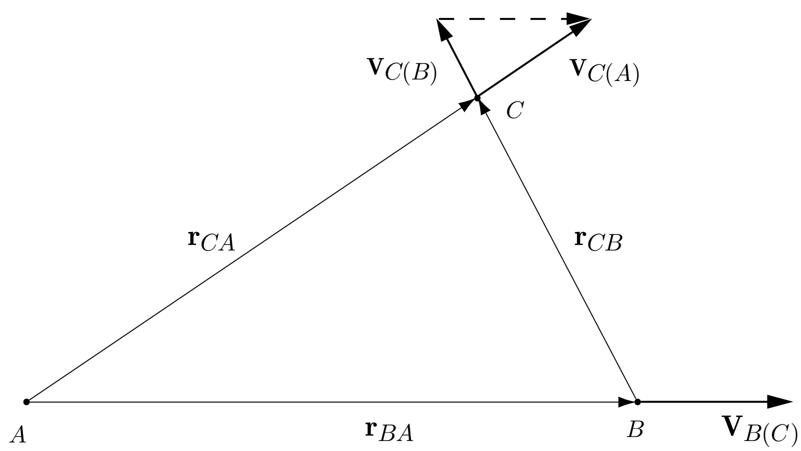

Fig. 1.1.

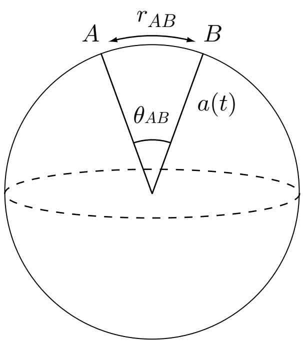

Fig. 1.2.

Problem 1.1 In order for a general expansion law, $\mathbf{v} = \mathbf{f}(\mathbf{r},t)$, to be the same for all observers, the function $\mathbf{f}$ must satisfy the relation

$$\mathbf{f}(\mathbf{r}_{CA} - \mathbf{r}_{BA},t) = \mathbf{f}(\mathbf{r}_{CA},t) - \mathbf{f}(\mathbf{r}_{BA}, t). \tag{1.4}$$

Show that the only solution of this equation is given by (1.1).

A useful analogy for envisioning Hubble expansion is the two-dimensional surface of an expanding sphere (Figure 1.2). The angle θ_{AB} between any two points A and B on the surface of the sphere remains unchanged as its radius $a(t)$ increases. Therefore the distance between the points, measured along the surface, grows as

$$r_{AB}(t) = a(t)\theta_{AB}, \tag{1.5}$$

implying a relative velocity

$$v_{AB} = \dot{r}_{AB} = \dot{a}(t)\theta_{AB} = \frac{\dot{a}}{a} r_{AB}, \tag{1.6}$$

where dot denotes a derivative with respect to time t. Thus, the Hubble law emerges here with $H(t) \equiv \dot{a}/a$.

The distance between any two observers A and B in a homogeneous and isotropic universe can be also rewritten in a form similar to (1.5). Integrating the equation

$$\dot{\mathbf{r}}_{BA} = H(t)\mathbf{r}_{BA}, \tag{1.7}$$

we obtain

$$\mathbf{r}_{BA}(t) = a(t)\chi_{BA}, \tag{1.8}$$

where

$$a(t) = \exp\left(\int H(t)\, dt\right) \tag{1.9}$$

is called the scale factor and is the analogue of the radius of the 2-sphere. The integration constant, χ_{BA}, is the analogue of θ_{BA} and can be interpreted as the distance between points A and B at some particular moment of time. It is called the Lagrangian or *comoving* coordinate of B, assuming a coordinate system centered at A.

In the 2-sphere analogy, $a(t)$ has a precise geometrical interpretation as the radius of the sphere and, consequently, has a fixed normalization. In Newtonian theory, however, the value of the scale factor $a(t)$ itself has no geometrical meaning and its normalization can be chosen arbitrarily. Once the normalization is fixed, the scale factor $a(t)$ describes the distance between observers as a function of time. For example, when the scale factor increases by a factor of 3, the distance between any two observers increases threefold. Therefore, when we say the size of the universe was, for instance, 1000 times smaller, this means that the distance between any two comoving objects was 1000 times smaller – a statement which makes sense even in an infinitely large universe. The Hubble parameter, which is equal to

$$H(t) = \frac{\dot{a}}{a}, \tag{1.10}$$

measures the expansion rate.

In this description, we are assuming a perfectly homogeneous and isotropic universe in which all observers are comoving in the sense that their coordinates χ remain unchanged. In the real universe, wherever matter is concentrated, the motion of nearby objects is dominated by the inhomogeneities in the gravitational field, which lead, for example, to virial orbital motion rather than Hubble expansion. Similarly, objects held together by other, stronger forces resist Hubble expansion. The velocity of these objects relative to comoving observers is referred to as the "peculiar" velocity. Hence, the Hubble law is valid only on the scales of homogeneity.

Problem 1.2 Typical peculiar velocities of galaxies are about a few hundred kilometers per second. The mean distance between large galaxies is about 1 Mpc. How distant must a galaxy be from us for its peculiar velocity to be small compared to its comoving (Hubble) velocity, if the Hubble parameter is 75 km s^{-1} Mpc^{-1}?

The current value of the Hubble parameter, H_0, can be determined by measuring the ratio of the recession velocity to the distance for an object whose peculiar velocity is small compared to its comoving velocity. The recessional velocity can be accurately measured because it induces a Doppler shift in spectral lines. The challenge is to find a reliable measure of the distance. Two methods used are based on the concepts of "standard candles" and "standard rulers." A class of objects is called a standard candle if the objects have about the same luminosity. Usually, they possess a set of characteristics that can be used to identify them even when they are far away. For example, Cepheid variable stars pulse at a periodic rate, and Type IA supernovae are bright, exploding stars with a characteristic spectral pattern. The distances to nearby objects in the class are measured directly (for example, by parallax) or by comparing them to another standard candle whose distance has already been calibrated. Once the distance to a subset of a given standard candle class has been measured, the distance to further members of that class can be determined: the inverse square law relates the apparent luminosity of the distant objects to that of the nearby objects whose distance is already determined. The standard ruler method is exactly like the standard candle method except that it relies on identifying a class of objects of the same size rather than the same luminosity. It is clear, however, that only if the variation in luminosity or size of objects within the same class is small can they be useful for measuring the Hubble parameter. Cepheid variable stars have been studied for nearly a century and appear to be good standard candles. Type IA supernovae are promising candidates which are potentially important because they can be observed at much greater distances than Cepheids. Because of systematic uncertainties, the value of the measured Hubble constant is known today with only modest accuracy and is about 65–80 km s^{-1} Mpc^{-1}.

Knowing the value of the Hubble constant, we can obtain a rough estimate for the age of the universe. If we neglect gravity and consider the velocity to be constant in time, then two points separated by $|\mathbf{r}|$ today, coincided in the past, $t_0 \simeq |\mathbf{r}|/|\mathbf{v}| = 1/H_0$ ago. For the measured value of the Hubble constant, t_0 is about 15 billion years. We will show later that the exact value for the age of the universe differs from this rough estimate by a factor of order unity, depending on the composition and curvature of the universe.

Because the Hubble law has a kinematical origin and its form is dictated by the requirement of homogeneity and isotropy, it has to be valid in both Newtonian theory and General Relativity. In fact, rewritten in the form (1.8), it can be immediately applied in Einstein's theory. This remark may be disconcerting since, according to the Hubble law, the relative velocity can exceed the speed of light for two objects separated by a distance larger than $1/H$. How can this be consistent with Special Relativity? The resolution of the paradox is that, in General Relativity, the relative velocity has no invariant meaning for objects whose separation exceeds $1/H$, which represents the curvature scale. We will explore this point further in context of the Milne universe (Section 1.3.5), following the discussion of Newtonian cosmology.

1.2 Dynamics of dust in Newtonian cosmology

We first consider an infinite, expanding, homogeneous and isotropic universe filled with "dust," a euphemism for matter whose pressure p is negligible compared to its energy density ε. (In cosmology the terms "dust" and "matter" are used interchangeably to represent nonrelativistic particles.) Let us choose some arbitrary point as the origin and consider an expanding sphere about that origin with radius $R(t) = a(t)\chi_{com}$. Provided that gravity is weak and the radius is small enough that the speed of the particles within the sphere relative to the origin is much less than the speed of light, the expansion can be described by Newtonian gravity. (Actually, General Relativity is involved here in an indirect way. We assume the net effect on a particle within the sphere due to the matter outside the sphere is zero, a premise that is ultimately justified by Birkhoff's theorem in General Relativity.)

1.2.1 Continuity equation

The total mass M within the sphere is conserved. Therefore, the energy density due to the mass of the particles is

$$\varepsilon(t) = \frac{M}{(4\pi/3)R^3(t)} = \varepsilon_0\left(\frac{a_0}{a(t)}\right)^3, \tag{1.11}$$

where ε_0 is the energy density at the moment when the scale factor is equal a_0. It is convenient to rewrite this conservation law in differential form. Taking the time derivative of (1.11), we obtain

$$\dot{\varepsilon}(t) = -3\varepsilon_0\left(\frac{a_0}{a(t)}\right)^3\frac{\dot{a}}{a} = -3H\varepsilon(t). \tag{1.12}$$

This equation is a particular case of the nonrelativistic continuity equation,

$$\frac{\partial\varepsilon}{\partial t} = -\nabla(\varepsilon\mathbf{v}), \tag{1.13}$$

if we take $\varepsilon(\mathbf{x}, t) = \varepsilon(t)$ and $\mathbf{v} = H(t)\mathbf{r}$. Beginning with the continuity equation and assuming homogeneous initial conditions, it is straightforward to show that the unique velocity distribution which maintains homogeneity evolving in time is the Hubble law: $\mathbf{v} = H(t)\mathbf{r}$.

1.2.2 Acceleration equation

Matter is gravitationally self-attractive and this causes the expansion of the universe to decelerate. To derive the equation of motion for the scale factor, consider a probe particle of mass m on the surface of the sphere, a distance $R(t)$ from the origin. Assuming matter outside the sphere does not exert a gravitational force on the particle, the only force acting is due to the mass M of all particles within the sphere. The equation of motion, therefore, is

$$m\ddot{R} = -\frac{GmM}{R^2} = -\frac{4\pi}{3}Gm\frac{M}{(4\pi/3)R^3}R. \tag{1.14}$$

Using the expression for the energy density in (1.11) and substituting $R(t) = a(t)\chi_{com}$, we obtain

$$\ddot{a} = -\frac{4\pi}{3}G\varepsilon a. \tag{1.15}$$

The mass of the probe particle and the comoving size of the sphere χ_{com} drop out of the final equation.

Equations (1.12) and (1.15) are the two master equations that determine the evolution of $a(t)$ and $\varepsilon(t)$. They *exactly* coincide with the corresponding equations for dust ($p = 0$) in General Relativity. This is not as surprising as it may seem at first. The equations derived do not depend on the size of the auxiliary sphere and, therefore, are exactly the same for an infinitesimally small sphere where all the particles move with infinitesimal velocities and create a negligible gravitational field. In this limit, General Relativity *exactly* reduces to Newtonian theory and, hence, relativistic corrections should not arise.

1.2.3 Newtonian solutions

The closed form equation for the scale factor is obtained by substituting the expression for the energy density (1.11) into the acceleration equation (1.15):

$$\ddot{a} = -\frac{4\pi}{3}G\varepsilon_0\frac{a_0^3}{a^2}. \tag{1.16}$$

Multiplying this equation by $\dot{a}$ and integrating, we find

$$\frac{1}{2}\dot{a}^2 + V(a) = E, \tag{1.17}$$

where E is a constant of integration and

$$V(a) = -\frac{4\pi G\varepsilon_0 a_0^3}{3a}.$$

Equation (1.17) is identical to the energy conservation equation for a rocket launched from the surface of the Earth with unit mass and speed $\dot{a}$. The integration constant E represents the total energy of the rocket. Escape from the Earth occurs if the positive kinetic energy overcomes the negative gravitational potential or, equivalently, if E is positive. If the kinetic energy is too small, the total energy E is negative and the rocket falls back to Earth. Similarly, the fate of the dust-dominated universe – whether it expands forever or eventually recollapses – depends on the sign of E. As pointed out above, the normalization of a has no invariant meaning in Newtonian gravity and it can be rescaled by an arbitrary factor. Hence, only the sign of E is physically relevant. Rewriting (1.17) as

$$H^2 - \frac{2E}{a^2} = \frac{8\pi G}{3}\varepsilon, \tag{1.18}$$

we see that the sign of E is determined by the relation between the Hubble parameter, which determines the kinetic energy of expansion, and the mass density, which defines the gravitational potential energy.

In the rocket problem, the mass of the Earth is given and the student is asked to compute the minimal escape velocity by setting $E = 0$ and solving for the velocity v. In cosmology, the expansion velocity, as set by the Hubble parameter, has been reasonably well measured while the mass density was very poorly determined for most of the twentieth century. For this historical reason, the boundary between escape and gravitational entrapment is traditionally characterized by a critical density, rather than critical velocity. Setting $E = 0$ in (1.18), we obtain

$$\varepsilon^{cr} = \frac{3H^2}{8\pi G}. \tag{1.19}$$

The critical density decreases with time since H is decreasing, though the term "critical density" is often used to refer to its current value. Expressing E in terms of the energy density $\varepsilon(t)$ and the Hubble constant $H(t)$, we find

$$E = \frac{4\pi G}{3}a^2\varepsilon^{cr}\left(1 - \frac{\varepsilon}{\varepsilon^{cr}}\right) = \frac{4\pi G}{3}a^2\varepsilon^{cr}\left[1 - \Omega(t)\right], \tag{1.20}$$

where

$$\Omega(t) \equiv \frac{\varepsilon(t)}{\varepsilon^{cr}(t)} \tag{1.21}$$

is called the cosmological parameter. Generally, $\Omega(t)$ varies with time, but because the sign of E is fixed, the difference $1 - \Omega(t)$ does not change sign. Therefore, by measuring the current value of the cosmological parameter, $\Omega_0 \equiv \Omega(t_0)$, we can determine the sign of E.

We shall see that the sign of E determines the spatial geometry of the universe in General Relativity. In particular, the spatial curvature has the opposite sign to E. Hence, *in a dust-dominated universe*, there is a direct link between the ratio of the energy density to the critical density, the spatial geometry and the future evolution of the universe. If $\Omega_0 = \varepsilon_0/\varepsilon_0^{cr} > 1$, then $E < 0$ and the spatial curvature is positive (closed universe). In this case the scale factor reaches some maximal value and the universe recollapses, as shown in Figure 1.3. When $\Omega_0 < 1$, E is positive, the spatial curvature is negative (open universe), and the universe expands hyperbolically. The