
Snippets of Physics

15. Hubble Expansion for Pedestrians

T Padmanabhan

Many features of the expanding universe, which should be legitimately discussed using general relativity, can be sneaked in by using Newtonian physics in an expanding coordinate system. In this special issue on Hubble, I describe several of these features along with some cautionary comments.

Since the large-scale dynamics of the universe is essentially governed by gravity, any theoretical model for gravity will have implications for the large-scale physics of the universe. This was known, of course, even to Newton who did attempt to describe the universe using his ideas of gravity. We now know, however, that the proper description of gravity should be based on Einstein's general relativity rather than on Newtonian ideas.

It was mentioned in a previous installment that the description of gravity in Einstein's theory is based on the notion of curved spacetime. In the case of the universe, this will involve treating it as a curved spacetime with a geometry determined by the distribution of matter. The simplest of such models treats the distribution of matter in the universe as uniform and isotropic (at sufficiently large scales) and tries to understand the properties of the curved spacetime produced by such a matter distribution. While all these might sound complicated, a surprising feature about such a model of the universe is that much of its dynamics can be understood fairly easily – without introducing complicated notions from general relativity. Needless to say, such an approach is filled with pitfalls and one needs to constantly verify that one is not getting carried away by the simplifying



T Padmanabhan works at IUCAA, Pune and is interested in all areas of theoretical physics, especially those which have something to do with gravity.

Keywords

Cosmology, expansion of the universe, structure formation.



It is not possible to introduce a set of coordinates on the surface of a sphere such that the 'line interval' reduces to the Pythagorean form. This is the difference between curved and flat space.

flavour of Newtonian physics. In this installment, I will describe how these ideas work.

The difference between a flat space and a curved space can be encoded in the generalization of Pythagoras theorem for infinitesimally separated points. For example, a flat 2-dimensional surface (say, a plain sheet of paper) allows us to introduce standard Cartesian coordinates (x, y) such that the distance between infinitesimally separated points can be expressed in the form $dl^2 = dx^2 + dy^2$ which, of course, is just the standard Pythagoras theorem. In contrast, consider the two-dimensional surface on a sphere of radius r on which we have introduced two angular coordinates (θ, ϕ) . The corresponding formula will now read $dl^2 = r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2$. It is not possible to introduce any other set of coordinates on the surface of a sphere such that this expression – usually called the ‘line interval’ – reduces to the Pythagorean form. This is the difference between a curved space and flat space.

Move on from space to spacetime and from points to events. In flat spacetime, which we use in special relativity, the ‘Pythagoras theorem’ generalizes to the form

$$ds^2 = -c^2 dt^2 + dx^2 + dy^2 + dz^2. \quad (1)$$

The spatial coordinates appear in the standard form and the inclusion of time introduces the all important minus sign. But one can live with it and treat it as a generalization of the formula $dl^2 = dx^2 + dy^2$ to 4-dimensions (with an extra minus sign). But in a curved spacetime, this expression will not hold and the coordinate differentials like $c^2 dt^2$, dx^2 , etc., in the interval will get multiplied by functions of space and time. This is just like our using $\sin^2 \theta d\phi^2$ rather than just $d\phi^2$ to describe the curved 2-dimensional surface of a sphere. The precise manner in which such a modification occurs is determined by Einstein’s equation and depends on the distribution of matter in spacetime.



While this can lead to pretty complicated spacetimes in general, the large-scale universe turns out to be described by a remarkably simple generalization of the line interval in (1). We only need to modify it into the form

$$ds^2 = -c^2 dt^2 + a^2(t) [dx^2 + dy^2 + dz^2], \quad (2)$$

where the function $a(t)$ is called the ‘expansion factor’. All the information about the behaviour of the universe is contained in this single function which – in turn – can be determined by Einstein’s equation if we know the contents of the universe¹. However, even without knowing the explicit form of $a(t)$, one can figure out a lot of things about such a universe, as we shall see.

The key trick is to notice that, at any given time t , one can introduce a new spatial coordinate $\mathbf{r}(t) \equiv a(t)\mathbf{x}$ so that, at this instant of time, the space looks just like what we are accustomed to in special relativity. The \mathbf{r} is called ‘proper coordinate’ while \mathbf{x} is called the ‘comoving coordinate’. Since the space looks familiar in terms of \mathbf{r} , one uses standard laws of physics in terms of these proper coordinates and then translates them back to \mathbf{x} , hoping for the best. Amazingly, it works for most purposes.

To begin with, consider two particles located at \mathbf{x}_1 and $\mathbf{x}_2 = \mathbf{x}_1 + \delta\mathbf{x}$ which are infinitesimally separated. The coordinate distance between these two particles is $|\delta\mathbf{x}|$ while the proper distance is $|\delta\mathbf{l}(t)| = a(t)|\delta\mathbf{x}|$. We now note that, even if the particles do not move in terms of \mathbf{x} -coordinate (i.e., each particle has a fixed \mathbf{x} -coordinate which does not change with time) their *proper* separation changes with time because of the $a(t)$ factor. The relative velocity at which these particles are moving from each other is given by

$$\delta\mathbf{v} = \frac{d\delta\mathbf{l}}{dt} = \dot{a}\delta\mathbf{x} = \frac{\dot{a}}{a}\delta\mathbf{l} . \quad (3)$$

This result, as we can see, is essentially Hubble’s law! It shows that the two particles are moving away from

¹As usual, we are simplifying the universe a little bit; it turns out that you actually need one more number characterizing the curvature of the space to describe the universe completely. But observations show that this number is quite close to zero and hence I will ignore it.



each other with a speed proportional to their separation when $H(t) \equiv (\dot{a}/a)$ is positive.

Given this result, one can obtain several other interesting consequences. Suppose a narrow pencil of (nearly) monochromatic electromagnetic radiation crosses these two comoving observers located at \mathbf{x}_1 and $\mathbf{x}_2 = \mathbf{x}_1 + \delta\mathbf{x}$. We want to know what frequency these two observers will attribute to the electromagnetic radiation. The time for the electromagnetic radiation to traverse the distance δl will be $\delta t = \delta l/c$. Let the frequency of the radiation measured by the first observer be ω . Since the first observer sees the second one to be *receding* with velocity δv , she will expect the second observer to measure a Doppler shifted frequency $(\omega + \delta\omega)$, where

$$\frac{\delta\omega}{\omega} = -\frac{\delta v}{c} = -\frac{\dot{a} \delta l}{a c} = -\frac{\dot{a}}{a} \delta t = -\frac{\delta a}{a} . \quad (4)$$

How does one interpret this relation? It shows that the frequency of electromagnetic radiation as measured by the comoving observers changes when the expansion factor $a(t)$ changes with time. If δa is positive (i.e., if the universe is expanding), $\delta\omega$ is negative indicating a redshift in the frequency of radiation. In fact, the above equation can be immediately integrated to give

$$\omega(t)a(t) = \text{constant} . \quad (5)$$

We thus conclude that the frequency of electromagnetic radiation changes due to expansion of the universe according to the law $\omega \propto a^{-1}$. This approach works because, in an infinitesimal region around an event, one can always use the laws of special relativity. (One can think of it as a variant of the so-called *principle of equivalence* which essentially tells you that the genuine effects of spacetime curvature are second order in the separation between close events.) This is true in any spacetime but we will not usually be able to integrate the local result and obtain a global law in a general spacetime when

The frequency of electromagnetic radiation changes due to expansion of the universe according to the law $\omega \propto a^{-1}$.



the geometry varies from point to point in space. We could achieve it in this particular case because the modification of spacetime interval from the one in (1) to the one in (2) did not involve any function that depended on the *spatial* coordinates.

In fact, one can obtain a similar result for any particle, not just photons. To do this, let us consider a material particle which passes the first observer with velocity v . When it has crossed the proper distance δl (in a time interval δt), it passes the second observer whose velocity (relative to the first one) is

$$\delta u = \frac{\dot{a}}{a} \delta l = \frac{\dot{a}}{a} v dt = v \frac{\delta a}{a} . \quad (6)$$

The velocity attributed to this particle by the second observer can be obtained by using the special relativistic law for the addition of velocities. This gives

$$\begin{aligned} v' &= \frac{v - \delta u}{1 - (v\delta u/c^2)} = v - \left(1 - \frac{v^2}{c^2}\right) \delta u + \mathcal{O} \left[\frac{(\delta u)^2}{c^2} \right] \\ &= v - \left(1 - \frac{v^2}{c^2}\right) v \frac{\delta a}{a} . \end{aligned} \quad (7)$$

Rewriting this equation as

$$\delta v = -v \left(1 - \frac{v^2}{c^2}\right) \frac{\delta a}{a} \quad (8)$$

and integrating, we get

$$p = \frac{v}{\sqrt{1 - (v^2/c^2)}} = \frac{\text{constant}}{a} . \quad (9)$$

In other words, the magnitude of the 3-momentum decreases as a^{-1} due to the expansion. If the particle is non-relativistic, then $v \propto p$ and velocity itself decays as a^{-1} .



Once we have the scaling of the momentum of particles and photons, one can proceed further and understand how the energy density of radiation changes in an expanding universe. To do this, let us consider the description of a bunch of photons in terms of a distribution function $f(\mathbf{r}, \mathbf{p}, t)$ in phase space. As usual, $dN = f(\mathbf{r}, \mathbf{p}, t)d^3\mathbf{r}d^3\mathbf{p}$ gives the number of photons in a small phase volume. In expanding coordinates, the spatial volume increases as a^3 while, from our previous result, we know that the momentum space volume decreases as a^{-3} . So the phase space volume element is invariant and – since dN is invariant – the distribution function f remains invariant as the universe expands. Expressing the momentum of the photon as $\mathbf{p} = (\hbar\omega/c)\hat{\mathbf{p}}$, where $\hat{\mathbf{p}}$ is a unit vector in the direction of propagation, we find that the momentum space volume is proportional to $\omega^2 d\omega d\Omega$ where $d\Omega$ denotes the solid angle in the direction of the momentum $\hat{\mathbf{p}}$. So we can also write

$$dN = f(\mathbf{r}, \mathbf{p}, t)d^3\mathbf{r}d^3\mathbf{p} \propto f(\mathbf{r}, \omega, \hat{\mathbf{p}}, t)\omega^2 d^3\mathbf{r} d\omega d\Omega \quad (10)$$

Further, because f gives the *number* density of photons per unit phase volume, the corresponding energy density is given by $(\hbar\omega)f$. It follows that the energy density of radiation per unit range of frequency is proportional to $\rho \propto \omega^3 f$. Using our previous result that the distribution function remains invariant, we conclude that $\rho(\omega)/\omega^3$ remains invariant as the universe expands.

Using our previous result that the distribution function remains invariant, we conclude that $\rho(\omega)/\omega^3$ remains invariant as the universe expands.

This has a very interesting consequence. Suppose the universe is filled with a radiation bath, the energy density of which has the form $\rho(\omega) = \omega^3 F(\omega/\alpha)$, where α is some parameter and F is some arbitrary function of its argument. As the universe expands, ρ/ω^3 remains invariant while ω itself changes as $\omega \propto a(t)^{-1}$. If we denote the values measured today by a subscript 0, then $\omega_0 = \omega(t)a(t)/a_0$. It follows that for the factor ω_0/α , we



can write

$$\frac{\omega_0}{\alpha} = \frac{\omega(t)(a(t)/a_0)}{\alpha} = \frac{\omega(t)}{\alpha(a_0/a(t))} . \quad (11)$$

This shows that the redshifting of the frequency can be equivalently thought of as rescaling of the parameter α as the universe expands with $\alpha(t) \propto 1/a(t)$. So the radiation energy density of the form $\rho(\omega) = \omega^3 F(\omega/\alpha)$ retains its shape as the universe expands, except for an overall scaling.

The Planck spectrum of radiation is one special case in which energy density has the above-mentioned functional form with

$$\rho \propto \omega^3 [\exp(\hbar\omega/kT) - 1]^{-1} \equiv \omega^3 F(\omega/T) . \quad (12)$$

The relevant parameter now is the temperature of the radiation. Therefore, as the universe expands, a Planck spectrum remains a Planck spectrum with the temperature redshifting according to the law $T \propto a^{-1}$.

It should be stressed that this result has nothing to do with thermal equilibrium! In fact, the situation we are considering is precisely the other extreme of thermal equilibrium in which the radiation has completely decoupled from matter. To clarify this point, let me briefly describe what happens in our real universe. Very early in the evolution of the universe, charged particles and photons were strongly coupled to each other and existed in the form of plasma in *real* thermodynamic equilibrium. Such a strong coupling implies that the radiation will be thermalized and its spectral distribution would have the Planckian form. Let us assume that at some instant of time, we switch off all the interaction between radiation and matter (in our universe this happened when it was about one-thousandth of the present size). From that epoch onwards, each of the photons has been propagating in the expanding universe with its frequency redshifting according to the law $\omega \propto a^{-1}$. The photons are

As the universe expands, a Planck spectrum remains a Planck spectrum with the temperature redshifting according to the law $T \propto a^{-1}$.



In Wheelerian language, we get “thermal equilibrium without thermal equilibrium”.

not interacting with matter: there is no exchange of energy and there is no process which ‘maintains’ thermal equilibrium. Nevertheless, at present, the photons will be described by a Planck distribution with a redshifted temperature, because the form of the Planck spectrum will be preserved with the temperature of the radiation decreasing as $T \propto a^{-1}$ under cosmic expansion. This is a purely kinematic effect occurring for photons which are propagating freely through the universe. In Wheelerian language, we get “thermal equilibrium without thermal equilibrium”.

Let us next take a closer look at the dynamics of non-relativistic particles in such an expanding universe. Consider a particle located at the comoving coordinate \mathbf{x} corresponding to the proper coordinate $\mathbf{r}(t) = a(t)\mathbf{x}$. Even if \mathbf{x} does not change with time – i.e., even if the particle has constant comoving coordinate – its proper coordinate \mathbf{r} will change with time due to $a(t)$. This will induce an acceleration on the particle given by $\ddot{\mathbf{r}} = (\ddot{a}/a)\mathbf{r}$. Given our usual prejudice that accelerations arise due to forces, it seems natural to attribute this acceleration to the existence of a global “cosmic potential” $\Phi = -(1/2)(\ddot{a}/a)r^2$, so that we can write $\ddot{\mathbf{r}} = -\nabla_r \Phi$. (The subscript r in ∇_r is to remind ourselves that the gradient is with respect to r and not x ; note that $\nabla_r = a^{-1}\nabla_x$.) Within the context of such Newtonian considerations, we can attribute this potential to a mass density $\rho_{\text{bg}}(t)$ such that $\nabla_r^2 \Phi = 4\pi G \rho_{\text{bg}}(t)$. Simple differentiation of Φ gives

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\rho_{\text{bg}}(t) \tag{13}$$

which relates the expansion factor $a(t)$ to a uniform background density $\rho_{\text{bg}}(t)$ of matter in the universe. It all seems natural to assume that the total number of particles within a proper volume should not change as the universe expands, thereby suggesting $\rho_{\text{bg}}(t) \propto a^{-3}$. If you substitute this into (13), then it is easy to show



that $a(t) \propto t^{2/3}$ even though we will not require this result in our discussion. In this approach, we think of a large number of particles being distributed uniformly throughout the universe and attribute Φ to such a collection of particles. This is, of course, conceptually a dubious procedure within the context of Newtonian gravity but it turns out that – for the specific case under discussion – general relativity leads to the same result.

Given this interpretation, one can ask what happens if we perturb the density in the universe in a space-dependent manner so that $\rho_{\text{bg}}(t) \rightarrow \rho_{\text{bg}}(t)[1 + \delta(t, \mathbf{x})]$. The potential will change to $\Psi \equiv \Phi + \phi$ with the extra bit ϕ produced by the extra density $\rho_{\text{bg}}(t)\delta(t, \mathbf{x})$; that is:

$$\nabla_r^2 \phi = \frac{1}{a^2} \nabla_x^2 \phi = 4\pi G \rho_{\text{bg}}(t) \delta(t, \mathbf{x}) \quad . \quad (14)$$

We would like to interpret this situation in terms of a system of a large number of particles with comoving positions \mathbf{x}_i , where $i = 1, 2, \dots$ labels each of the particles. As long as all the particles have constant values for \mathbf{x}_i , all of them will move away from each other due to the expansion which we attribute to the potential Φ . This is the spatially uniform density situation. But if the particles are disturbed from their constant \mathbf{x}_i values, then the matter density will become nonuniform – with $\delta(t, \mathbf{x}) \neq 0$ – and the potential Φ is modified to $\Psi \equiv \Phi + \phi$. Of course, each of the particles will now feel the force due to this total potential Ψ . The acceleration of the j -th particle, now given by

$$\frac{d^2 \mathbf{r}_j}{dt^2} = \ddot{a} \mathbf{x}_j + 2\dot{a} \dot{\mathbf{x}}_j + a \ddot{\mathbf{x}}_j \quad (15)$$

arises due to the gradient of the modified potential $\Psi \equiv \Phi + \phi$. We note that the first term in the right-hand side of (15) is $(\ddot{a}/a)\mathbf{r}$ which is just $-\nabla_r \Phi$. Therefore, $\nabla_r \phi (= a^{-1} \nabla_x \phi)$ should lead to the other two terms in



Equation (16) is the key to understanding gravitational clustering in an expanding background.

(15). With some rearrangement, this leads to

$$\ddot{\mathbf{x}}_j + 2\frac{\dot{a}}{a}\dot{\mathbf{x}}_j = -\frac{1}{a^2}\nabla_x\phi \quad (16)$$

This equation tells you how the comoving coordinates of the particles change thereby making the density distribution of particles in the universe non-uniform which, in turn, gives rise to $\nabla_x\phi$. The potential ϕ can be thought of as being generated by the perturbations from the uniform density of particles. This equation is the key to understanding gravitational clustering in an expanding background.

Each of the terms in this equation has an interesting interpretation. The first term $\ddot{\mathbf{x}}_j$ is the acceleration corresponding to the comoving position of the particle, which arises over and above the acceleration due to the background expansion (the first term in (15) which we have already accounted for by the gradient of cosmic potential). The second term is a damping (friction) term which tries to decrease the speed of the particle. In fact, if the right-hand side of (16) is zero, the equation can be integrated to give $a^2\dot{\mathbf{x}} = \text{constant}$ or, alternatively $a\dot{\mathbf{x}} = 1/a$. This can be rewritten in the form

$$a\dot{\mathbf{x}} = \dot{\mathbf{r}} - H\mathbf{r} = \frac{1}{a} \quad (17)$$

The left-hand side is the deviation of the proper velocity $\dot{\mathbf{r}}$ from the Hubble expansion velocity $H\dot{\mathbf{r}}$. The quantity $a\dot{\mathbf{x}}$ is sometimes called ‘peculiar velocity’ for no good reason. The right-hand side shows that this difference decays down a $1/a$ so that – in the absence of the $\nabla_x\phi$ term on the right hand side of (16) – particles will tend to approach the cosmic expansion velocity. This is the key effect of the ‘friction’ term.

Finally, the gradient in the right hand side of (16) is what is keeping the acceleration alive and leads to gravitational clustering. Here too, there is one curious feature. The perturbation of the background density, $\delta(t, \mathbf{x})$,



can be either positive or negative in any given region (except for the condition that $\delta > -1$ to keep $\rho > 0$). Therefore, the source for the potential ϕ in (14) can be positive density or negative density! (This is somewhat like electrostatics in which the charge density can be positive or negative.) So the gravitational force produced by this distribution can be attractive or repulsive in the right-hand side of (16)! In fact an underdense region of the universe – usually called a ‘void’ in the distribution of galaxies – exerts an effective repulsive force on the surrounding matter.

An underdense region of the universe – usually called a ‘void’ in the distribution of galaxies – exerts an effective repulsive force on the surrounding matter.

By multiplying (16) throughout by a^2 we can recast it in the form

$$\frac{d}{dt}(a^2 \dot{\mathbf{x}}_j) = -\nabla_x \phi \quad . \quad (18)$$

Obviously, this equation for each of the particles can be obtained from an explicitly time-dependent Lagrangian of the form

$$L = \sum_j \left(\frac{1}{2} m a^2 \dot{x}_j^2 - m \phi \right) \equiv K - U, \quad (19)$$

where we have summed over all the particles with the understanding that ϕ at the location of each particle is produced by the rest of the particles. This Lagrangian, in turn, will lead to a Hamiltonian of the form

$$H = \sum_j \left(\frac{p_j^2}{2ma^2} + m\phi \right) \equiv K + U. \quad (20)$$

This allows us to obtain a rather interesting and curious result. We first recall that whenever a Hamiltonian depends explicitly on time, we have the result $(dH/dt) = (\partial H/\partial t)$ where the right-hand side is evaluated keeping the coordinates and momenta constant. From the form of H , we immediately get $(\partial K/\partial t) = -2(\dot{a}/a)K$. On the other hand, the potential energy between any two particles varies as $|\mathbf{r}_i - \mathbf{r}_j|^{-1} = a^{-1}|\mathbf{x}_i - \mathbf{x}_j|^{-1}$. This implies



that $(\partial U/\partial t) = -(\dot{a}/a)U$. Putting all these together, we get the result

$$\frac{d}{dt}(K + U) = -\frac{\dot{a}}{a}(2K + U). \quad (21)$$

This relation goes under the name ‘Cosmic Virial theorem’ (or ‘Cosmic Energy equation’) and is one of the few exact results you can obtain rather easily in this particular context. The left-hand side of the equation represents the rate of change of total energy of a collection of particles. The right-hand side tells you that $(K + U)$ is not conserved for such a system of particles except in two different contexts. The first one is the rather trivial case of $\dot{a} = 0$ which is just standard classical mechanics without any background expansion and the total energy is, of course, conserved. The second – and more curious situation – corresponds to $2K + U = 0$ which you will recognize is the standard virial equilibrium condition for a set of particles interacting via Newtonian gravity. In practical terms, this result implies the following. Suppose during the evolution of the universe a bunch of particles come together and form a virialized self-gravitating cluster. Then, to the extent we can ignore the interaction of this cluster with the rest of the particles in the universe, its energy will be conserved. Roughly speaking, such virialized clusters do not participate in the cosmic dynamics.

Finally I want to describe an interesting and exact bound on the kinetic energy of particles in such a cluster formed in the expanding universe, which can be obtained from the cosmic virial theorem. To do this, we first note that (21) can also be written as:

$$\frac{d}{dt}a(K + U) = -K\dot{a} < 0 \quad (22)$$

So we know that $a(K + U)$ is a decreasing function of time. Very early on, when no significant clustering has

Virialized clusters
do not participate
in the cosmic
dynamics.



occurred, we will have $K \approx 0$ and $U < 0$, making $(K + U)$ negative very early in the evolution. It then follows that we must have $K < -U$ or rather $K < |U|$ at all times. Next, note that (21) can also be written as

$$\frac{d}{dt} a^2 \left(K + \frac{1}{2} U \right) = - \frac{a^2}{2} \frac{dU}{dt} . \quad (23)$$

As structures develop in the universe, potential wells will get deeper and deeper and hence $d(-U)/dt > 0$ making the left-hand side positive. In the early stages, since $U \approx 0$, we have $(K + U/2) > 0$. Hence we conclude that, at any later time $K > -(1/2)U$ or $K > (1/2)|U|$. Combining with the previous result, we get

$$\frac{1}{2}|U| < K < |U| . \quad (24)$$

This is a rather neat result on the relationship between kinetic and potential energies of structures formed by gravitational clustering, which must hold independent of the details of the process!

Suggested Reading

- [1] P J E Peebles, *Large scale structure of the universe*, Princeton University Press, 1980, section 24 .
- [2] T Padmanabhan, *Structure formation in the universe*, Cambridge University Press, 1992, Chapter 4.

Address for Correspondence

T Padmanabhan
 IUCAA, Post Bag 4
 Pune University Campus
 Ganeshkhind
 Pune 411 007, India.
 Email:
 paddy@iucaa.ernet.in
 nabhan@iucaa.ernet.in

