Cosmology Part I: The Unperturbed Universe

Hiranya V. Peiris[1,*]
[1]*Institute of Astronomy, University of Cambridge, Cambridge, CB1 3AX, U.K.*

Classes

M.W.F. 11 am, *MR9*

Office

Obs 28, Old Observatory Building, Institute of Astronomy, Madingley Road (Tel: 337533).

Course Website

`http://camd05.ast.cam.ac.uk/Cosmology/`

Introductory Reading:

1. Liddle, A. An Introduction to Modern Cosmology. Wiley (2003)

Complementary Reading

1. Dodelson, S. Modern Cosmology. Academic Press (2003) $*$
2. Carroll, S.M. Spacetime and Geometry. Addison-Wesley (2004) $*$
3. Liddle, A.R. and Lyth, D.H. Cosmological Inflation and Large-Scale Structure. Cambridge (2000) $*$
4. Kolb, E.W. and Turner, M.S. The Early Universe. Addison-Wesley (1990) $*$
5. Weinberg, S. Gravitation and Cosmology. Wiley (1972) $*$
6. Peacock, J.A. Cosmological Physics. Cambridge (2000)
7. Mukhanov, V. Physical Foundations of Cosmology. Cambridge (2005)

Books denoted with a $*$ are particularly recommended for this course.

Acknowledgements

This course is distilled from a variety of excellent presentations of the subject in courses and textbooks I have encountered over the years. It is, of course, a topic that can be treated with a wide range of sophistication and difficulty. I have elected to aim it at an accessible level with more emphasis on developing physical intuition rather than on mathematical principles. In terms of textbooks, it most closely parallels the treatments found in Dodelson and Carroll, which are my current personal favourites, and borrows with gratitude from course notes by (in no particular order) Paul Steinhardt, Paul Shellard, Daniel Baumann, and Anthony Lasenby. I thank Anthony Challinor for reading through a huge sheaf of rough notes for this course and providing helpful comments. I am supported by STFC and the European Commission.

Errata

Any errata contained in the following notes are solely my own. Reports of any typos or unclear explanations in the notes will be gratefully received at the email address below. The notes are evolving, and the most up-to-date version at any given time will be found on the website above.

*Electronic address: `hiranya@ast.cam.ac.uk`

Contents

I. INTRODUCTION 4
- A. Brief history of the universe 4
- B. The universe observed 5

II. THE METRIC 7
- A. The cosmological principle 7
- B. Example metrics 7
- C. The metric and the Einstein summation convention 9
- D. Special relativity metric 9
- E. General relativity metric 10
- F. Metric of a spatially-flat expanding universe 11

III. THE GEODESIC EQUATION 12
- A. Transforming a Cartesian basis 12
- B. Christoffel symbol 13
- C. Particles in an expanding universe 14
- D. Redshift 15

IV. THE EINSTEIN EQUATION 16
- A. Energy-momentum tensor 16
  1. Dust 17
  2. Perfect fluid 17
  3. Evolution of energy 17
- B. Friedmann equations in a flat universe 19

V. GENERAL FRIEDMANN-ROBERTSON-WALKER METRIC 20
- A. The cosmological principle revisited 20
  1. Maximally symmetric spaces 20
  2. Spatial curvature 21
  3. General FRW metric 22
  4. General Friedmann equations 23
- B. Dynamics of the FRW universe 24
  1. Terminology 24
  2. Evolution of the scale factor 25
- C. Matter-radiation equality 26
- D. Cosmological constant 26
- E. Spatial curvature and destiny 27

VI. TIME AND DISTANCE 30
- A. Conformal time 30
- B. Lookback distance and lookback time 31
- C. Instantaneous physical distance 32
- D. Luminosity distance 32
- E. Angular diameter distance 33

VII. PARTICLES AND FIELDS IN COSMOLOGY 35
- A. General particle motion 35
- B. Classical field theory 35
- C. Energy-momentum tensor from the action 37

VIII. BASICS OF INFLATION 41
- A. Shortcomings of the Standard Big Bang Theory 41
  1. Homogeneity Problem 41
  2. Flatness Problem 41
  3. Horizon Problem 42
  4. On the Problem of Initial Conditions 43
  5. What got the Big Bang going? 43
- B. Possible resolution of the Big Bang puzzles 43
  1. A Crucial Idea 43
  2. Comoving Horizon during Inflation 44
  3. Flatness Problem Revisited 45
  4. Horizon Problem Revisited 46
  5. Origin of primordial inhomogeneities 46
  6. Conformal Diagram of Inflationary Cosmology 46

IX. INFLATION: IMPLEMENTATION WITH A SCALAR FIELD 48
- A. Negative pressure 48
- B. Implementation with a scalar field 48
- C. Slow-Roll Inflation 50

D. What is the Physics of Inflation? 51
E. Slow-Roll Inflation in the Hamilton-Jacobi Approach (NON–EXAMINABLE) 52
1. Hamilton-Jacobi Formalism 52
2. Hubble Slow-Roll Parameters 52
3. Slow-Roll Inflation 53

I. INTRODUCTION

A. Brief history of the universe

The discovery of the expansion of the universe by Edwin Hubble in 1929 heralded the dawn of observational cosmology. If we mentally rewind the expansion, we find that the universe was hotter and denser in its past. At very early times the temperature was high enough to ionize the material that filled the universe. The universe therefore consisted of a plasma of nuclei, electrons and photons, and the number density of free electrons was so high that the mean free path for the Thomson scattering of photons was extremely short. As the universe expanded, it cooled, and the mean photon energy diminished. The universe transitioned from being dominated by radiation to being matter-dominated. Eventually, at a temperature of about 3000° K, the photon energies became too low to keep the universe ionized. At this time, known as *recombination*, the primordial plasma coalesced into neutral atoms, and the mean free path of the photons increased to roughly the size of the observable universe. Initial inhomogeneities present in the primordial plasma grew under the action of gravitational instability during the matter-dominated era into all the bound structures we observe in the universe today. Now, 13.7 billion years later, it appears that the universe has entered an epoch of accelerated expansion, with its energy density dominated by the mysterious “dark energy”.

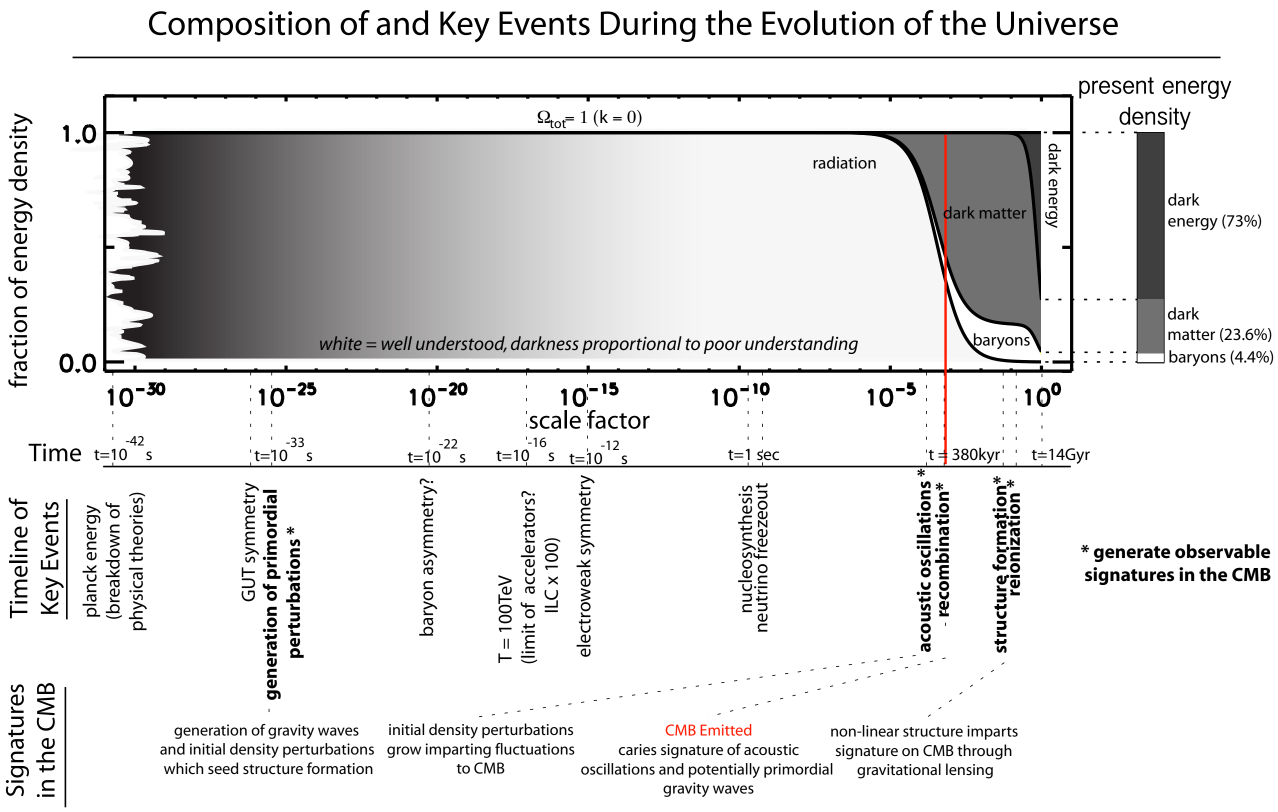

FIG. 1 Composition of, and key events during, the history of the universe. Figure credit: Jeff McMahon.

The history of the universe is summarized in Fig. 1, emphasizing the “known unknowns” in our understanding of its composition and evolution. Bear in mind that there might be “unknown unknowns” as well!

The following table [reproduced from Liddle and Lyth] summarizes key events in the history of the universe and the corresponding time– and energy–scales:

t	$\rho^{1/4}$	Event
10^{-42} s	10^{18} GeV	Inflation begins?
$10^{-32\pm6}$ s	$10^{13\pm3}$ GeV	Inflation ends, Cold Big Bang begins?
$10^{-18\pm6}$ s	$10^{6\pm3}$ GeV	Hot Big Bang begins?
10^{-10} s	100 GeV	Electroweak phase transition?
10^{-4} s	100 MeV	Quark-hadron phase transition?
10^{-2} s	10 MeV	γ, ν, e, $\bar{e}$, n, and p in thermal equilibrium
1 s	1 MeV	ν decoupling, $e\bar{e}$ annihilation.
100 s	0.1 MeV	Nucleosynthesis (BBN)
10^4 yr	1 eV	Matter-radiation equality
10^5 yr	0.1 eV	Atom formation, photon decoupling (CMB)
$\sim 10^9$ yr	10^{-3} eV	First bound structures form
Now	10^{-4} eV (2.73 K)	The present.

During most of its history, the universe is very well described by the *hot Big Bang* theory – *i.e.* the idea that the universe was hot and dense in the past and has since cooled by expansion. The observational pillars underlying the Big Bang Theory are:

- the Hubble diagram,
- Big Bang Nucleosynthesis (BBN),
- the Cosmic Microwave Background (CMB).

In this course, amongst (many) other things, we will explore the theoretical underpinnings of our understanding of these observations. Cosmological observations have indicated several properties of the universe that cannot be explained within the "standard model" (both the hot Big Bang theory and the Standard Model of particle physics):

- dark matter,
- dark energy,
- anisotropies of the CMB.

We will treat the first two phenomenologically in this course, and develop a basic understanding of the third.

B. The universe observed

The observed universe has the following properties:

1. homogeneous and isotropic when averaged over the largest scales (homogeneous means isotropic from every vantage point).
2. topologically trivial (not periodic within 3000 Mpc).
3. expanding (Hubble flow).
4. hotter in the past, cooling (CMB).
5. predominantly matter (*vs.* antimatter).
6. chemical composition (75% H, 25% He, trace "metals": hot Big Bang plus processing in stars).
7. highly inhomogeneous today and locally ($\ll 100$ Mpc).

8. highly homogeneous when it was $10^3\times$ smaller.
9. (maybe) negligible curvature.
10. (maybe) dark energy/acceleration.

In these lectures, $\boxed{\text{we will use natural units, } \hbar = c = k_B = 1}$, unless explicitly knowing the dependence on these quantities is necessary to develop understanding.

Most of cosmology can be learnt with only a passing knowledge of general relativity (GR). We will need the concepts of the *metric* and the *geodesic*, and apply *Einstein's equations* to the Friedmann-Robertson-Walker metric, relating the metric parameters to the (energy) density of the universe. In this section of the course, we will apply Einstein's equations to the *unperturbed* universe. In the third section of the course, we will apply them to the *perturbed* universe. With the experience we gain here, there will be nothing difficult later. The principles are identical; only the algebra will be a bit harder.

II. THE METRIC

A. The cosmological principle

On the largest scales, the universe is assumed to be uniform. This idea is called the *cosmological principle.* There are two aspects of the cosmological principle:

- the universe is *homogeneous.* There is no preferred observing position in the universe.
- the universe is *isotropic.* The universe looks the same in every direction.

Fig. 2 illustrates these concepts. We will make them more precise in due course.

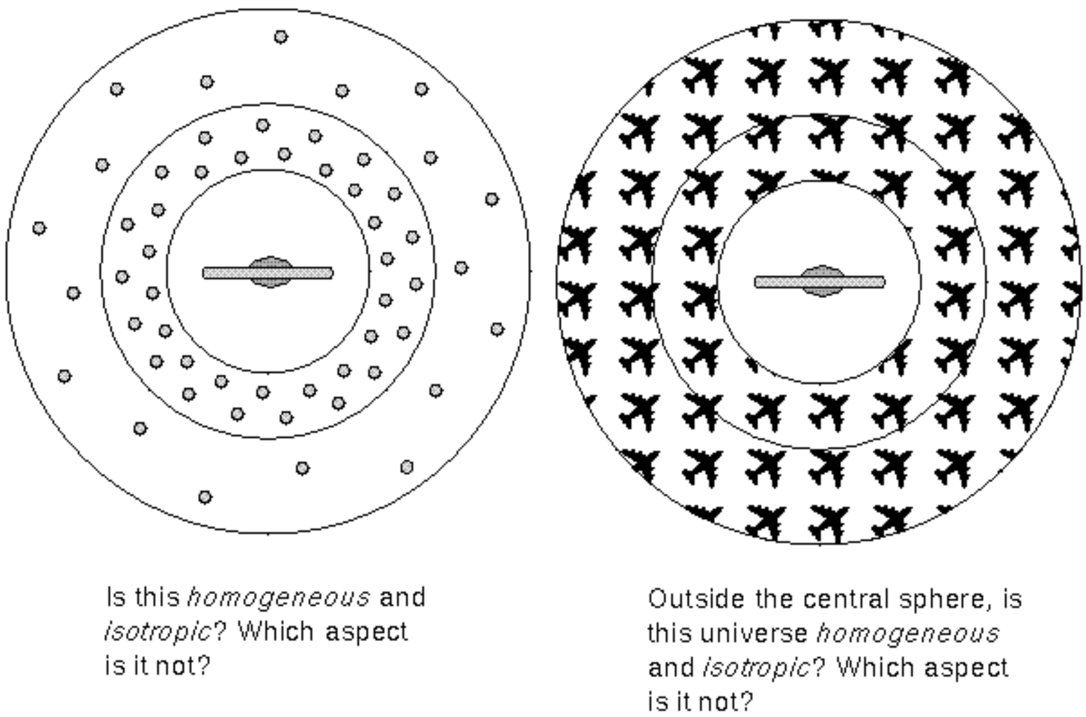

FIG. 2 Departures from homogeneity and isotropy illustrated. Figure credit: This image was produced by Nick Strobel, and obtained from *Nick Strobel's Astronomy Notes* at `http://www.astronomynotes.com`.

There is an overwhelming amount of observational evidence that the universe is *expanding.* This means that early in the history of the universe, the distant galaxies were closer to us than they are today. It is convenient to describe the scaling of the coordinate grid in an expanding universe by the *scale factor.* In a smooth, expanding universe, the scale factor connects the coordinate distance with the physical distance. More generally,

$$\text{coordinate distance} \Longrightarrow \boxed{\text{metric}} \Longrightarrow \text{physical distance}\,. \tag{1.2.1}$$

The metric is an essential tool to make quantitative predictions in an expanding universe.

B. Example metrics

Cartesian coordinates:

$$ds^2 = dx^2 + dy^2. \tag{1.2.2}$$

Polar coordinates:

$$ds^2 = dr^2 + r^2 d\theta^2 \neq dr^2 + d\theta^2. \tag{1.2.3}$$

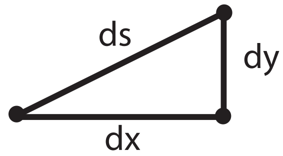

FIG. 3 2D Cartesian Coordinates.

A metric turns observer-dependent coordinates into invariants. In 2D (Fig. 3),

$$ds^2 = \sum_{i,j=1,2} g_{ij} dx^i dx^j \ , \tag{1.2.4}$$

where the metric g_{ij} is a 2×2 symmetric matrix. In Cartesian coordinates,

$$x^1 = x, \quad x^2 = y, \quad g_{ij} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \ , \tag{1.2.5}$$

while in polar coordinates,

$$x^1 = r, \quad x^2 = \theta, \quad g_{ij} = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix} \ . \tag{1.2.6}$$

Another way to think about a metric is to take a pair of vectors on a topographical map (Fig. 4), of the same length (same coordinate distance). But the actual *physical* distance depends on the topography $\leftrightarrow$ metric.

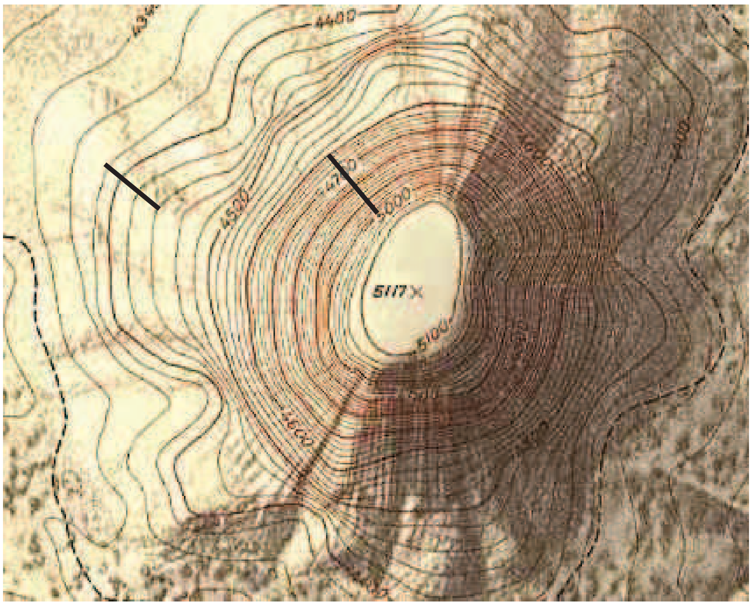

FIG. 4 Contour map of a mountain, with closely spaced contours near the centre corresponding to rapid elevation gain.

The great advantage of a metric is that it incorporates gravity. In classical Newtonian mechanics, gravity is an external force, and particles move in a gravitational field. In GR, gravity is encoded in the metric, and the particles move in a distorted/curved spacetime. In 4 (3+1) dimensions, the invariant includes *time intervals* as well:

$$ds^2 = \sum_{\mu,\nu=0}^{3} g_{\mu\nu} dx^\mu dx^\nu, \tag{1.2.7}$$

where $\mu, \nu \longrightarrow \{0, 1, 2, 3\}$ with $dx^0 = dt$ reserved for the *timelike* coordinate, and dx^i for the *spacelike* coordinates. From now, we will use the summation convention where repeated indices are summed over.

C. The metric and the Einstein summation convention

In 3D, a vector $\vec{A}$ has three components: $A^i, \quad i = \{1,2,3\}$.

$$\Rightarrow \vec{A}\cdot\vec{B} = \sum_{i=1}^{3} A^i B^i \equiv A^i B^i \ , \tag{1.2.8}$$

where we sum over repeated indices. For example, the matrix product: $\mathbf{M}\cdot\mathbf{N} = (\mathbf{M}\cdot\mathbf{N})_{ij} = M_{ik}N_{kj}$, where summation over k is implied.

In relativity, there are two generalizations. First, there is a zeroth component, time. Spatial indices run from $1 \to 3$, and it is conventional to use 0 for the time component: $A^\mu = (A^0, A^i)$. Secondly, there is a distinction between the upper (vectors) and lower (1–forms) indices. One goes back and forth between the two using the metric:

$$A_\mu = g_{\mu\nu}A^\nu; \quad A^\mu = g^{\mu\nu}A_\nu. \tag{1.2.9}$$

A vector and a 1–form can be *contracted* to produce an invariant; a scalar. For example, the statement "The squared 4–momentum of a massless particle must vanish." is equivalent to:

$$p^2 \equiv p_\mu p^\mu = g_{\mu\nu}p^\mu p^\nu = 0. \tag{1.2.10}$$

Contraction can be thought of as counting the contours of the topographic map crossed by a vector, in our previous analogy. The metric can be used to raise and lower indices on tensors with an arbitrary number of indices. For example,

$$g^{\mu\nu} = g^{\mu\alpha}g^{\nu\beta}g_{\alpha\beta}. \tag{1.2.11}$$

Taking $\alpha = \nu$, we see that

$$\boxed{g^{\nu\beta}g_{\alpha\beta} = \delta^\nu{}_\alpha} \ , \tag{1.2.12}$$

where $\delta^\nu{}_\alpha$ is the Kronecker delta:

$$\delta^\nu{}_\alpha = \begin{cases} 1 & (\nu = \alpha) \\ 0 & (\nu \neq \alpha) \end{cases} \ . \tag{1.2.13}$$

This is the definition of $g^{\mu\nu}$ as the inverse of $g_{\mu\nu}$. The metric $g_{\mu\nu}$ is

- necessarily symmetric,
- in principle, has 4 diagonal and 6 off-diagonal components,
- provides the connection between the values of the coordinates and the more physical measure of the *interval* ds^2.

In this course we will use the following convention for the *signature* of the metric: $(+.-,-,-)$. Beware, while this convention is commonly used by particle physicists, the convention used in relativity and cosmology is often $(-,+,+,+)$.

D. Special relativity metric

The *Minkowski metric* $\eta_{\mu\nu}$ is the metric of special relativity, and it describes flat space. Its line element is given by,

$$ds^2 = \eta_{\mu\nu}dx^\mu dx^\nu \ , \tag{1.2.14}$$
$$= dt^2 - \left(dx^2 + dy^2 + dz^2\right) \ , \tag{1.2.15}$$

and the metric is

$$\boxed{\eta_{\mu\nu} = \mathrm{diag}(1,-1,-1,-1)} \ . \tag{1.2.16}$$

SR applies in inertial frames, or locally, in those falling freely in a gravitational field. Because it is locally equivalent only, we can't say in general that there is a frame where

$$\Delta s^2 = \Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2 = 0 \ . \tag{1.2.17}$$

We want to be able to transform to non-free-fall coordinate systems and relate variables over distances. This will require a more general set of transformations than the Lorentz transformations (translations, rotations, and boosts). There is thus *no global SR frame*, and $g_{\mu\nu}$ has to reflect the curvature of spacetime.

E. General relativity metric

Let's consider two forms of a key principle.

Strong Equivalence Principle (SEP): At any point in a gravitational field, in a frame moving with the free fall acceleration at that point, *all the laws of physics* have their usual Special Relativity (SR) form, except *gravity*, which disappears locally.

Weak Equivalence Principle (WEP): At any point in a gravitational field, in a frame moving with the free fall acceleration at that point, *the laws of motion of free test particles* have their usual Special Relativity (SR) form, *i.e.* particles move in straight lines with uniform velocity locally.

Instead of thinking of particles moving under a force, the WEP allows us to think of them in a frame *without* gravity, but moving with the free fall acceleration at that point. This is a very powerful idea:

$$\text{EQUIVALENCE}: \quad \boxed{\text{gravity} \Longleftrightarrow \text{acceleration}}\,, \tag{1.2.18}$$

which has two important consequences. First, it explains the equivalence of gravitational mass and inertial mass. Second, it says that the motion of a test body in a gravitational field *only* depends on its *position* and *instantaneous velocity* in spacetime. In other words, $\boxed{\text{gravity determines a geometry}}$. The equivalence principle works *locally* where the gravitational field can be taken to be *uniform*.

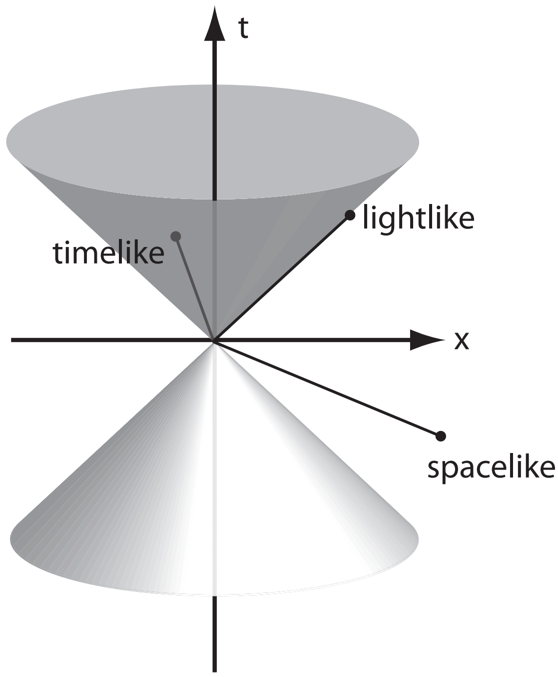

FIG. 5 A lightcone on a spacetime diagram. Points that are spacelike, lightlike, and timelike separated from the origin are indicated.

There is a component of "true" gravity, not transformable into acceleration, that has the distance dependence of "tidal forces" ($\propto r^{-3}$ for spherical polar coordinates). True gravity manifests itself via the separation or coming together of test particles initially on parallel tracks. There is no *global Lorentz frame* in the presence of a *non-uniform* gravitational field.

A freely-falling particle follows a *geodesic* in *spacetime*. The *metric* links the concepts of "geodesic" and "spacetime":

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu \tag{1.2.19}$$

where ds^2 is the proper interval, $g_{\mu\nu}$ is the metric tensor, and x^μ is a four-vector. There are three possible kinds of intervals (see Fig. 5):

$$ds^2 < 0 : \quad \text{spacelike} \tag{1.2.20}$$
$$ds^2 = 0 : \quad \text{null / lightlike} \tag{1.2.21}$$
$$ds^2 > 0 : \quad \text{timelike} \; . \tag{1.2.22}$$

F. Metric of a spatially-flat expanding universe

What is the metric of an *expanding universe*?

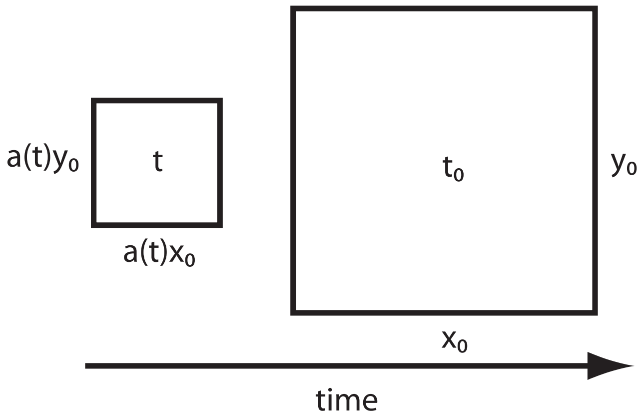

FIG. 6 If the comoving distance today at time t_0 is x_0, the physical distance between the two points at some earlier time $t < t_0$ was $a(t)x_0$.

If the *comoving distance* today is x_0, the *physical distance* between two points at some earlier time t was $a(t)x_0$ (see Fig. 6). At least in a *flat* (as opposed to open or closed) universe, the metric must be $\sim$ Minkowski, except that the distance must be multiplied by the scale factor $a(t)$. Thus, the metric of a flat, expanding universe is the Friedmann-Robertson-Walker metric:

$$\boxed{g_{\mu\nu} = \text{diag}\left(1, -a^2(t), -a^2(t), -a^2(t)\right)} \; . \tag{1.2.23}$$

The evolution of the scale factor depends on the density of the universe. When perturbations are introduced, the metric will become more complicated, and the perturbed part of the metric will become determined by the inhomogeneities in the matter and the radiation.

III. THE GEODESIC EQUATION

A. Transforming a Cartesian basis

In Minkowski space, particles travel in straight lines unless they are acted upon by an external force. In more general spacetimes, the concept of a "straight line" gets replaced by the "geodesic", which is the path followed by a particle in the absence of any external forces. Let us generalize Newton's law with no forces, $\frac{d^2\vec{x}}{dt^2}=0$, to the expanding universe. We will start with particle motion in a Euclidean 2D plane. Equations of motion in Cartesian coordinates $x^i=(x,y)$ for a free particle are

$$\frac{d^2x^i}{dt^2}=0\ . \tag{1.3.1}$$

What are the equations of motion in polar coordinates, ${x'}^i=(r,\theta)$? The basis vectors for polar coordinates, $\hat{r}$, $\hat{\theta}$ vary in the plane! Therefore, in polar coordinates,

$$\frac{d^2\vec{x}}{dt^2}=0 \quad \not\Rightarrow \quad \frac{d^2{x'}^i}{dt^2}=0 \tag{1.3.2}$$

for ${x'}^i=(r,\theta)$. Let us start from the Cartesian equation and transform:

$$\frac{dx^i}{dt}=\left(\frac{dx^i}{dx'^j}\right)\frac{dx'^j}{dt}\ , \tag{1.3.3}$$

where the term in the brackets on the RHS is a *transformation matrix* going from one basis to another, *i.e.* it is the determinant of the Jacobean. To transform from Cartesian to polar coordinates,

$$x^1=x'^1\cos x'^2, \quad x^2=x'^1\sin x'^2\ , \tag{1.3.4}$$

with transformation matrix,

$$\frac{dx^i}{dx'^j}=\begin{pmatrix}\cos x'^2 & -x'^1\sin x'^2\\ \sin x'^2 & x'^1\cos x'^2\end{pmatrix}\ . \tag{1.3.5}$$

Therefore, the geodesic equation is,

$$\frac{d}{dt}\left[\frac{dx^i}{dt}\right]=\frac{d}{dt}\left[\frac{dx^i}{dx'^j}\frac{dx'^j}{dt}\right]=0\ . \tag{1.3.6}$$

If the transformation was *linear*, the derivative acting on the transformation matrix would vanish, and the geodesic equation in the new basis would still be $\frac{d^2{x'}^i}{dt^2}=0$. In polar coordinates, the transformation is *not linear*, and using the chain rule, we have

$$\frac{d}{dt}\left[\frac{dx^i}{dx'^j}\right]=\frac{dx'^k}{dt}\frac{\partial^2x^i}{\partial x'^k\partial x'^j}\ . \tag{1.3.7}$$

The geodesic equation therefore becomes,

$$\frac{d}{dt}\left[\frac{dx^i}{dx'^j}\frac{dx'^j}{dt}\right]=\frac{\partial x^i}{\partial x'^j}\frac{d^2x'^j}{dt^2}+\frac{\partial^2x^i}{\partial x'^j\partial x'^k}\frac{dx'^k}{dt}\frac{dx'^j}{dt}\ . \tag{1.3.8}$$

Multiplying by the inverse of the transformation matrix, we obtain the geodesic equation in a non-Cartesian basis:

$$\frac{d^2x'^\ell}{dt^2}+\left[\left(\left\{\frac{\partial x}{\partial x'}\right\}^{-1}\right)^\ell_{\ i}\frac{\partial^2x^i}{\partial x'^j\partial x'^k}\right]\frac{dx'^k}{dt}\frac{dx'^j}{dt}=0\ . \tag{1.3.9}$$

`EXERCISE:` Check that it works for polar coordinates!

The term in the brackets is the *Christoffel symbol*,

$$\Gamma^{\ell}_{jk} = \left[\left(\left\{ \frac{\partial x}{\partial x'} \right\}^{-1} \right)^{\ell}_{i} \frac{\partial^2 x^i}{\partial x'^j \partial x'^k} \right] , \tag{1.3.10}$$

which is symmetric in j, k. In Cartesian coordinates, $\Gamma^{\ell}_{jk} = 0$, and the geodesic equation is simply $\frac{d^2 x^i}{dt^2} = 0$. In general, $\Gamma^{\ell}_{jk} \neq 0$ describes geodesics in non-trivial coordinate systems. The geodesic equation is a very powerful concept, because in a non-trivial spacetime (*e.g.* the expanding universe) it is *not possible* to find a fixed Cartesian coordinate system. So we need to know how particles travel in the more general case.

To import this concept into relativity, we need

- to allow indices to range from $0 \to 3$ to include time and space.
- time is now one of our coordinates! We can't use it to describe an evolution parameter.

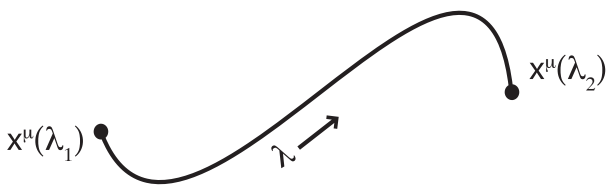

FIG. 7 A particle's path parametrized by λ, which monotonically increases from its initial value λ_1 to its final value λ_2.

Take a parameter λ (Fig. 7) which monotonically increases along the particle's path. The geodesic equation becomes:

$$\boxed{\frac{d^2 x^\mu}{d\lambda^2} + \Gamma^{\mu}_{\alpha\beta} \frac{dx^\alpha}{d\lambda} \frac{dx^\beta}{d\lambda} = 0} . \tag{1.3.11}$$

B. Christoffel symbol

Rather than the previous definition of the Christoffel symbol obtained by transforming a Cartesian basis, it is almost always more convenient to obtain it from the metric:

$$\boxed{\Gamma^{\mu}_{\alpha\beta} = \frac{g^{\mu\nu}}{2} \left[\frac{\partial g_{\alpha\nu}}{\partial x^\beta} + \frac{\partial g_{\beta\nu}}{\partial x^\alpha} - \frac{\partial g_{\alpha\beta}}{\partial x^\nu} \right]} . \tag{1.3.12}$$

`EXERCISE:` Verify that Eq. (1.3.10) is consistent with the definition (1.3.12) in the case of a flat spacetime.

Be careful: raised indices are important. $g^{\mu\nu}$ is the inverse of $g_{\mu\nu}$. So $g^{\mu\nu}$ in the flat FRW metric is identical to $g_{\mu\nu}$, except that the spatial elements are $-\frac{1}{a^2}$ instead of $-a^2$.

The components of the Christoffel symbol in the flat FRW universe (with overdots denoting d/dt) are:

$$\Gamma^0_{00} = 0, \quad \Gamma^0_{0i} = \Gamma^0_{i0} = 0, \quad \Gamma^0_{ij} = \delta_{ij}\dot{a}a \quad \Gamma^i_{0j} = \Gamma^i_{j0} = \frac{\dot{a}}{a}\,\delta^i{}_j, \quad \Gamma^i_{jk} = 0, \quad \Gamma^i_{00} = 0\ . \tag{1.3.13}$$

`EXERCISE:` Use the flat FRW metric to derive the components of the Christoffel symbol.

C. Particles in an expanding universe

Let us apply the geodesic equation to a single particle. How does a particle's energy change as the universe expands? Most measurements we make in cosmology has to do with intercepting photons which have arrived on the earth after being emitted at various epochs during the evolution of the universe. Therefore we will consider a massless particle, which has energy-momentum 4–vector, $p^\alpha = (E, \vec{p})$, and use this to implicitly define the parameter λ:

$$p^\alpha = \frac{dx^\alpha}{d\lambda}. \tag{1.3.14}$$

Eliminate λ by noting that $\frac{d}{d\lambda} = \frac{dx^0}{d\lambda}\frac{d}{dx^0} = E\frac{d}{dt}$, and the 0–component of the geodesic equation,

$$\begin{aligned} E\frac{dE}{dt} &= -\Gamma^0_{ij}p^i p^j\ , \\ &= -\delta_{ij}\dot{a}a\ p^i p^j\ . \end{aligned} \tag{1.3.15}$$

For a massless particle, the energy-momentum vector has zero magnitude:

$$g_{\mu\nu}p^\mu p^\nu = E^2 - \delta_{ij}a^2 p^i p^j = 0\ . \tag{1.3.16}$$

Since $\vec{p}$ measures motion on the comoving grid, the physical momentum which measures changes in the physical distance is related to $\vec{p}$ by a factor of a, hence the factor of a^2 here. This leads to

$$\frac{dE}{dt} + \frac{\dot{a}}{a}E = 0\ . \tag{1.3.17}$$

We see that the energy of a massless particle *decreases* as the universe expands:

$$\boxed{E \propto \frac{1}{a}}\ . \tag{1.3.18}$$

This accords with the intuition from a handwaving argument: $E \propto \lambda^{-1}$ (where λ is the wavelength) and $\lambda \propto a$ is stretched along with the expansion. The frequency of a photon emitted with frequency $\nu_{\rm em}$ will therefore be observed with a lower frequency $\nu_{\rm obs}$ as the universe expands:

$$\frac{\nu_{\rm obs}}{\nu_{\rm em}} = \frac{a_{\rm em}}{a_{\rm obs}}\ . \tag{1.3.19}$$

D. Redshift

Cosmologists like to speak of this in terms of the *redshift* z between two events, defined by the fractional change in wavelength:

$$z_{\mathrm{em}} = \frac{\lambda_{\mathrm{obs}} - \lambda_{\mathrm{em}}}{\lambda_{\mathrm{em}}} \,. \tag{1.3.20}$$

So if the observation takes place today ($a_{\mathrm{obs}} = a_0 = 1$), this implies

$$\boxed{a_{\mathrm{em}} = \frac{1}{1 + z_{\mathrm{em}}}} \,. \tag{1.3.21}$$

So the redshift of an object tells us the scale factor when the photon was emitted.

Notice that this redshift is not the same as the conventional Doppler effect. It is the expansion of space, not the relative velocities of the observer and emitter, that leads to the redshift.

Measuring the redshifts of distant objects is one of the most basic tools in the observer's toolkit. It is a rather amazing notion that every time one measures a redshift, one is directly detecting the curvature of spacetime.

IV. THE EINSTEIN EQUATION

So far, we have not used *General Relativity*. The concept of the metric and the realization that non-trivial metrics affect geodesics exist independently of GR. The part of GR that is hidden here is that *gravitation can be described by a metric*. There is another aspect of GR which we will need now, which connects the metric to the matter/energy. This is described by the Einstein equation, which relates *geometry* to *energy*:

$$\boxed{G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi G T_{\mu\nu}} \, , \tag{1.4.1}$$

where $G_{\mu\nu}$ is the Einstein tensor, $R_{\mu\nu}$ is the Ricci tensor, the Ricci scalar $R = g^{\mu\nu} R_{\mu\nu}$ is the contraction of the Ricci tensor, and the energy-momentum (or stress-energy) tensor $T_{\mu\nu}$ is a symmetric tensor describing the constituents of the universe. The Ricci tensor is defined as,

$$\boxed{R_{\mu\nu} = \Gamma^{\alpha}_{\mu\nu,\alpha} - \Gamma^{\alpha}_{\mu\alpha,\nu} + \Gamma^{\alpha}_{\beta\alpha}\Gamma^{\beta}_{\mu\nu} - \Gamma^{\alpha}_{\beta\nu}\Gamma^{\beta}_{\mu\alpha}} \, , \tag{1.4.2}$$

where commas denote derivatives with respect to x,

$$\Gamma^{\alpha}_{\mu\nu,\alpha} \equiv \frac{\partial \Gamma^{\alpha}_{\mu\nu}}{\partial x^{\alpha}} \, . \tag{1.4.3}$$

It looks like hard work but we have already done the hard bit by computing $\Gamma^{\alpha}_{\mu\nu}$ in a flat FRW universe, and it has only two sets of non-vanishing components: $\mu, \nu = 0$ and $\mu, \nu = i$. Using them (and noting that $\delta^{i}{}_{i} = 3$) we can show that,

$$R_{00} = -3 \left(\frac{\ddot{a}}{a} \right) , \quad R_{ij} = \delta_{ij} \left[2\dot{a}^2 + \ddot{a} a \right] \, . \tag{1.4.4}$$

Contracting the Ricci tensor, we then obtain the Ricci scalar for the flat, homogeneous FRW universe as:

$$\begin{aligned} R &\equiv g^{\mu\nu} R_{\mu\nu} \\ &= R_{00} - \frac{1}{a^2} \delta^{ij} R_{ij} \\ &= -6 \left[\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a} \right)^2 \right] \, . \end{aligned} \tag{1.4.5}$$

`EXERCISE:` Verify the components of the Ricci tensor and Ricci scalar given above for the flat, homogeneous FRW universe.

A. Energy-momentum tensor

Consider the 4-momentum (should be familiar from SR) of a particle: $p^{\mu} = mU^{\alpha}$, where m is the "rest mass" of the particle, independent of inertial frame. The energy of the particle is $E = p^0$ (the timelike component of the momentum 4-vector). Note that E is not invariant under Lorentz transformations. In the particle's rest frame, $p^0 = m$, which is just the famous result $E = mc^2$. We also have $p_{\mu} p^{\mu} = m^2$, or $E = \sqrt{m^2 + |\vec{p}|^2}$, where $\vec{p}^2 = \delta_{ij} p^i p^j$. p^{μ} provides a complete description of the energy-momentum of a *particle*. In cosmology, we need to describe extended systems comprised of large numbers of particles: *fluids*. A fluid is a continuum characterised by macroscopic properties such as density, pressure, entropy, and viscosity.

A single momentum 4-vector field is insufficient to describe the energy-momentum of a fluid. We need to replace it by the *energy-momentum tensor* (also called the stress-energy tensor) $T^{\mu\nu}$, which is a symmetric tensor describing *the flux of* 4*-momentum* p^{μ} *across a surface of constant* x^{ν}.

This definition is not that useful. Later, in the section on inflation, we will define $T^{\mu\nu}$ in terms of a functional derivative of the action with respect to the metric. This leads to a more algorithmic procedure for finding an explicit expression for $T^{\mu\nu}$. For now, we will use the above definition to gain some physical intuition.

Consider an infinitesimal element of the fluid in its rest frame, where there are no bulk motions. Then,

- T^{00}: the flux of p^0 (energy) in the x^0 (time) direction (*i.e.* the rest frame energy density ρ),
- $T^{0i} = T^{i0}$: momentum density,
- $T^{ij} = T^{ji}$: momentum flux (or *stress*) (*i.e.* forces between neighbouring infinitesimal elements of the fluid).

Off-diagonal elements in T^{ij} are shearing terms, *e.g.* due to viscosity. Diagonal terms T^{ii} give the i–th component of the force/unit area exerted by a fluid element in the i–direction. The pressure has three components given in the fluid rest frame. Let us now consider some example fluids.

1. Dust

Start with "dust" (cosmologists tend to use "matter" as a synonym for dust): in a flat spacetime, a collection of particles at rest with eachother. The number-flux 4-vector is $N^\mu = nU^\mu$, where n is the number density measured in the rest frame. N^0 is the number density of the particles in any other frame. N^i is the particle flux in the x^i direction. Imagine that each particle has mass m. The rest frame energy density is $\rho = mn$. ρ completely specifies dust, since pressure is by definition zero, as dust has no random motions within the fluid. Since $N^\mu = (n, 0, 0, 0)$, $p^\mu = (m, 0, 0, 0)$, we have

$$T^{\mu\nu}_{\text{dust}} = p^\mu N^\nu = mn\, U^\mu U^\nu = \rho\, U^\mu U^\nu \ . \tag{1.4.6}$$

This is not general enough to describe all interesting cosmological fluids, so we need to make a slight generalization.

2. Perfect fluid

A perfect fluid can be completely defined by a rest frame energy density ρ and an isotropic rest frame pressure P (which serves to specify pressure in every direction). Thus, $T^{\mu\nu}$ is diagonal in its rest frame, with no net flux of any component of momentum in an orthogonal direction. The non-zero spacelike components are all equal. For dust, we had $T^{\mu\nu} = \rho\, U^\mu U^\nu$. For a perfect fluid, the general form in any frame is:

$$\boxed{T^{\mu\nu} = (\rho + P)U^\mu U^\nu - P g^{\mu\nu}} \ . \tag{1.4.7}$$

`EXERCISE:` Check this by writing out the terms in *e.g.* flat space, $\eta_{\mu\nu}$.

We might have seemed to arrived at this arbitrarily, but given that (1.4.7) reduces to $T^{\mu\nu} = \text{diag}(\rho, -g^{ii}P)$ in the rest frame, and is a tensor equation (and therefore coordinate–independent), it must be valid in *any* frame. A perfect fluid is general enough to describe a wide variety of cosmological fluids, given their *equation of state*,

$$\boxed{w = \frac{P}{\rho}} \ . \tag{1.4.8}$$

Dust has $P = 0$, $w = 0$. Radiation has $P = \frac{\rho}{3}$, $w = \frac{1}{3}$. Vacuum energy is proportional to the metric: $T^{\mu\nu} = -\rho_{\text{vac}} g^{\mu\nu}$, $P_{\text{vac}} = -\rho_{\text{vac}}$, $w = -1$.

3. Evolution of energy

Consider a perfect isotropic fluid. Then, the energy-momentum tensor can be written with one index raised in the following metric-independent form:

$$T^\mu_{\ \nu} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & -P & 0 & 0 \\ 0 & 0 & -P & 0 \\ 0 & 0 & 0 & -P \end{pmatrix} , \tag{1.4.9}$$

where ρ is the energy density, and P is the pressure of the fluid. How do the components of $T^{\mu}_{\ \nu}$ evolve with time? Consider the case where there is no gravity and velocities are negligible. Then, the pressure and energy evolve as:

$$\text{Continuity Equation :} \qquad \frac{\partial \rho}{\partial t} = 0 \ , \tag{1.4.10}$$

$$\text{Euler Equation :} \qquad \frac{\partial P}{\partial x^i} = 0 \ . \tag{1.4.11}$$

We need to promote this to a 4–component conservation equation for the energy-momentum tensor:

$$\frac{\partial T^{\mu}_{\ \nu}}{\partial x^{\mu}} = 0. \tag{1.4.12}$$

However, in an expanding universe, the conservation criterion must be modified. In this context, conservation implies the vanishing of the *covariant derivative*:

$$\boxed{T^{\mu}_{\ \nu;\mu} \equiv \frac{\partial T^{\mu}_{\ \nu}}{\partial x^{\mu}} + \Gamma^{\mu}_{\alpha\mu} T^{\alpha}_{\ \nu} - \Gamma^{\alpha}_{\nu\mu} T^{\mu}_{\ \alpha}} \ . \tag{1.4.13}$$

The importance of the covariant derivative is that, to paraphrase Misner, Thorne, & Wheeler, *Gravitation* (W.H. Freeman, 1973) p. 387,

> A consistent replacement of regular partial derivatives by covariant derivatives carries the laws of physics (in component form) from flat (Lorentzian) spacetime into the curved (non-Lorentzian) spacetime of general relativity. Indeed, this substitution may be taken as a mathematical statement of Einstein's principle of equivalence.

We will call this the "comma goes to semi-colon" rule. Thus, the conservation criterion becomes

$$\boxed{T^{\mu}_{\ \nu;\mu} = 0} \ . \tag{1.4.14}$$

There are four separate equations to be considered here. Consider first the $\nu = 0$ component. By isotropy, $T^{i}_{\ 0}$ vanishes, yielding the *continuity equation*:

$$\boxed{\frac{\partial \rho}{\partial t} + 3\frac{\dot{a}}{a}(\rho + P) = 0} \ . \tag{1.4.15}$$

`EXERCISE:` Verify the above assertion.

Rearranging this equation,

$$a^{-3}\frac{\partial}{\partial t}\left(\rho a^3\right) = -3\left(\frac{\dot{a}}{a}\right)P \ . \tag{1.4.16}$$

Immediately this yields information about the scaling of both matter and radiation. Pressureless matter has $P = 0$ by definition, so

$$\text{MATTER :} \quad \frac{\partial}{\partial t}\left(\rho_{\mathrm{m}} a^3\right) = 0 \Rightarrow \boxed{\rho_{\mathrm{m}} \propto a^{-3}} \ . \tag{1.4.17}$$

This is expected if you consider that mass remains constant while number density scales as inverse volume.

Radiation has $P = \frac{\rho}{3}$, giving

$$\frac{\partial \rho_{\mathrm{r}}}{\partial t} + 4\left(\frac{\dot{a}}{a}\right)\rho_{\mathrm{r}} = a^{-4}\frac{\partial}{\partial t}\left(\rho_{\mathrm{r}} a^4\right) = 0 \ . \tag{1.4.18}$$

Thus, radiation scales as

$$\text{RADIATION :} \quad \boxed{\rho_{\mathrm{r}} \propto a^{-4}} \ . \tag{1.4.19}$$

`EXERCISE:` Show that $P = \frac{\rho}{3}$ for radiation.

B. Friedmann equations in a flat universe

To understand the evolution of the scale factor in a homogeneous expanding universe, we only need to consider the time-time component of the Einstein equation:

$$R_{00} - \frac{1}{2} g_{00} R = 8\pi G T_{00} \ , \tag{1.4.20}$$

leading to the first *Friedmann equation* for a flat universe:

$$\boxed{\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho} \quad \text{(FLAT)} \ . \tag{1.4.21}$$

`EXERCISE:` Verify the above assertion.

There is a second Friedmann equation. Consider the space-space component of Einstein's equation:

$$R_{ij} - \frac{1}{2} g_{ij} R = 8\pi G T_{ij} \ . \tag{1.4.22}$$

Using the flat FRW terms we worked out previously in Eqs. (1.3.13, 1.4.4, 1.4.5), with $g_{ij} = -\delta_{ij} a^2$ we find:

$$\text{LHS}: \quad \delta_{ij} \left[2\dot{a}^2 + \ddot{a}a\right] - \delta_{ij} \frac{a^2}{2}\, 6 \left[\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2\right] \ . \tag{1.4.23}$$

Noting the mixed form for the perfect fluid energy-momentum tensor, (1.4.9), we see that,

$$\text{RHS}: \quad 8\pi G T_{ij} = 8\pi G g_{ik} T^k_{\ j} = 8\pi G a^2 \delta_{ij} P \ . \tag{1.4.24}$$

Equating these terms we obtain,

$$\frac{\ddot{a}}{a} + \frac{1}{2}\left(\frac{\dot{a}}{a}\right)^2 = -4\pi G P \quad \text{(FLAT)} \ . \tag{1.4.25}$$

Combining with the first Friedmann equation (1.4.21), this leads us to the second *Friedmann equation*:

$$\boxed{\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P) \quad \text{(FLAT)}} \ . \tag{1.4.26}$$

`EXERCISE:` Derive Eq. (1.4.26) by another route, by differentiating the first Friedmann equation (1.4.21) with respect to time t, and combining with the continuity equation (1.4.15).

V. GENERAL FRIEDMANN-ROBERTSON-WALKER METRIC

A. The cosmological principle revisited

1. Maximally symmetric spaces

We leave the details to the concurrent GR course and/or any GR text, and note that the *Riemann tensor*:

$$R^{\rho}{}_{\sigma\mu\nu} = \partial_\mu \Gamma^{\rho}_{\nu\sigma} - \partial_\nu \Gamma^{\rho}_{\mu\sigma} + \Gamma^{\rho}_{\mu\lambda}\Gamma^{\lambda}_{\nu\sigma} - \Gamma^{\rho}_{\nu\lambda}\Gamma^{\lambda}_{\mu\sigma} \ , \tag{1.5.1}$$

$$R_{\rho\sigma\mu\nu} = g_{\rho\lambda} R^{\lambda}{}_{\sigma\mu\nu} \ , \tag{1.5.2}$$

quantifies *curvature* and is non-zero when the metric departs from *flatness*. The Ricci tensor is formed by contracting the Riemann tensor:

$$R_{\mu\nu} = g^{\alpha\sigma} R_{\alpha\mu\sigma\nu} = R^{\alpha}{}_{\mu\alpha\nu} \ . \tag{1.5.3}$$

It has some elegant symmetry properties, satisfying the following index symmetries:

$$R_{\alpha\mu\nu\sigma} = -R_{\mu\alpha\nu\sigma} = -R_{\alpha\mu\sigma\nu} \ , \quad R_{\alpha\mu\nu\sigma} = R_{\nu\sigma\alpha\mu} \ , \quad R_{\alpha\mu\nu\sigma} + R_{\alpha\sigma\mu\nu} + R_{\alpha\nu\sigma\mu} = 0 \ , \tag{1.5.4}$$

and the *Bianchi identities*,

$$R^{\lambda}{}_{\alpha\mu\nu;\sigma} + R^{\lambda}{}_{\alpha\sigma\mu;\nu} + R^{\lambda}{}_{\alpha\nu\sigma;\mu} = 0 \ . \tag{1.5.5}$$

In particular, in a *maximally symmetric space* (details left to GR course) of n dimensions,

$$R_{\rho\sigma\mu\nu} = \frac{R}{n(n-1)} \left(g_{\rho\mu} g_{\sigma\nu} - g_{\rho\nu} g_{\sigma\mu} \right) \ , \tag{1.5.6}$$

where the Ricci scalar R is constant over the manifold. Conversely, if the Riemann tensor satisfies (1.5.6), the space is maximally symmetric. Setting

$$K = \frac{R}{n(n-1)} \ , \tag{1.5.7}$$

since at any given point the metric can be put into its canonical form $g_{\mu\nu} = \eta_{\mu\nu}$, the kinds of maximally symmetric manifolds are characterized locally by the *metric signature* and the *sign* of K. For the metric signature $(+,-,-,-)$,

$$K = 0 : \quad \text{Minkowski space} \tag{1.5.8}$$

$$K < 0 : \quad \text{de Sitter space} \tag{1.5.9}$$

$$K > 0 : \quad \text{anti-de Sitter space} \ . \tag{1.5.10}$$

We said "locally" to account for possible global differences, such as between the plane and the torus.

Do any of these describe the real universe? Let's consider its properties. Contemporary cosmological models are based on the idea that, at "sufficiently large scales", the *Copernican principle* applies: the universe is pretty much the same everywhere. This is encoded more rigorously in the ideas of,

- *isotropy*: at some specified point in the manifold, space looks the same in whatever direction you look.
- *homogeneity*: the metric is the same throughout the manifold.

A manifold can be homogeneous but nowhere isotropic, or isotropic around a point but nowhere homogeneous. If a space is isotropic *everywhere*, then it is also homogeneous. If a space is isotropic around one point and also homogeneous, it will be isotropic everywhere.

The CMB shows that the universe is isotropic on the order of 10^{-5}, and since by the Copernican principle, we don't believe that we are the centre of the universe, we assume both homogeneity and isotropy.

However, observations tell us that the universe is *expanding*, so the Copernican principle only applies in space, not in time. So the maximally symmetric spacetimes itemized above don't describe our universe (or any universe with a dynamically interesting amount of matter and/or radiation).

2. Spatial curvature

So let's give up the "perfect" Copernican principle and posit that the universe is *spatially* homogeneous and isotropic:

$$ds^2 = dt^2 - R^2(t) d\sigma^2 \ , \tag{1.5.11}$$

where t is a timelike coordinate, $R(t)$ is the scale factor, and $d\sigma^2$ is the metric on a maximally symmetric 3–manifold Σ:

$$d\sigma^2 = \gamma_{ij}(u) du^i du^j \ , \tag{1.5.12}$$

where (u^1, u^2, u^3) are coordinates on Σ, and γ_{ij} is a maximally symmetric 3D metric. $R(t)$ tells us how big the spacelike slice is at time t. These are *comoving coordinates.*

We want to know the possible maximally symmetric 3-metrics γ_{ij}. They obey

$${}^{(3)}R_{ijkl} = k\left(\gamma_{ik}\gamma_{jl} - \gamma_{il}\gamma_{jk}\right) \ , \tag{1.5.13}$$

where

$$k = \frac{{}^{(3)}R}{6} \ , \tag{1.5.14}$$

and ${}^{(3)}$ reminds us that we are dealing with 3-metric γ_{ij}, not the entire spacetime metric. The Ricci tensor is then `(EXERCISE: CHECK!)`

$${}^{(3)}R_{jl} = 2k\gamma_{jl} \ . \tag{1.5.15}$$

Maximally symmetric $\Rightarrow$ spherically symmetric, so the metric can be put in the form (*cf.* Schwarzschild metric):

$$d\sigma^2 = \gamma_{ij} du^i du^j = e^{2\beta(\tilde{r})} d\tilde{r}^2 + \tilde{r}^2 d\Omega^2 \ , \tag{1.5.16}$$

where $\tilde{r}$ is the radial coordinate, and $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$ is the metric on the 2-sphere. Working out the components leads to:

$$\begin{aligned} {}^{(3)}R_{11} &= \frac{2}{\tilde{r}}\partial_1\beta \ , && (1.5.17)\\ {}^{(3)}R_{22} &= e^{-2\beta}\left(\tilde{r}\partial_1\beta - 1\right) + 1 \ , && (1.5.18)\\ {}^{(3)}R_{33} &= \left[e^{-2\beta}\left(\tilde{r}\partial_1\beta - 1\right) + 1\right]\sin\theta \ . && (1.5.19) \end{aligned}$$

Setting $\propto$ the metric via (1.5.15) , we get

$$\beta = -\frac{1}{2}\ln\left(1 - k\tilde{r}^2\right) \ . \tag{1.5.20}$$

Thus the metric on Σ is:

$$\boxed{d\sigma^2 = \frac{d\tilde{r}^2}{1 - k\tilde{r}^2} + \tilde{r}^2 d\Omega^2} \ . \tag{1.5.21}$$

The value of k sets the curvature, and therefore the size, of the spatial surfaces. It is common to normalize such that $k \in \{-1, 0, +1\}$, and absorb the physical size of the manifold into $R(t)$. The geometry is then classified as:

$$\begin{aligned} k = -1 :& \quad \text{constant negative curvature on } \Sigma \text{ (OPEN)} && (1.5.22)\\ k = 0 :& \quad \text{no curvature on } \Sigma \text{ (FLAT)} && (1.5.23)\\ k = +1 :& \quad \text{constant positive curvature on } \Sigma \text{ (CLOSED)} \ . && (1.5.24) \end{aligned}$$

The physical meaning of these cases becomes more apparent by redefining the radial coordinate:

$$d\chi = \frac{d\tilde{r}}{\sqrt{1 - k\tilde{r}^2}} \ . \tag{1.5.25}$$

Integrating,

$$\tilde{r} = S_k(\chi) \ , \tag{1.5.26}$$

where

$$S_k(\chi) \equiv \begin{cases} \sin\chi & k=+1 \\ \chi & k=0 \\ \sinh\chi & k=-1 \end{cases} \ , \tag{1.5.27}$$

such that

$$d\sigma^2 = d\chi^2 + S_k^2(\chi) d\Omega^2 \ . \tag{1.5.28}$$

`EXERCISE:` Verify Eq. (1.5.26)–(1.5.28).

To summarize, we have:

$$k=0: \qquad d\sigma^2 = d\chi^2 + \chi^2 d\Omega^2 = dx^2 + dy^2 + dz^2 \quad \text{(flat Euclidean space)} \tag{1.5.29}$$

$$k=+1: \qquad d\sigma^2 = d\chi^2 + \sin^2\chi \, d\Omega^2 \quad \text{(metric of a 3-sphere)} \tag{1.5.30}$$

$$k=-1: \qquad d\sigma^2 = d\chi^2 + \sinh^2\chi \, d\Omega^2 \quad \text{(a hyperboloid space)} \ . \tag{1.5.31}$$

A note on the hyperboloid case: globally, such a space could be infinite – the origin of "open" – but could also describe a non-simply-connected compact space, so it is not really a good description.

3. General FRW metric

The metric on spacetime describes one of these maximally symmetric hypersurfaces evolving in size:

$$ds^2 = dt^2 - R^2(t)\left[\frac{d\tilde{r}^2}{1-k\tilde{r}^2} + \tilde{r}^2 d\Omega^2\right] \ . \tag{1.5.32}$$

Normalizing the coordinates to the present epoch, subscript "0",

$$a(t) = \frac{R(t)}{R_0}, \quad r = R_0\tilde{r} \ , \tag{1.5.33}$$

we can define a curvature parameter of dimensions [length]$^{-2}$:

$$\kappa = \frac{k}{R_0}. \tag{1.5.34}$$

Note that κ can take any value, not just $\{+1, 0, -1\}$. We obtain the *general FRW metric*:

$$\boxed{ds^2 = dt^2 - a^2(t)\left[\frac{dr^2}{1-\kappa r^2} + r^2 d\Omega^2\right]} \ . \tag{1.5.35}$$

Setting $\dot{a} = \frac{da}{dt}$, the Christoffel symbols are:

$$\begin{array}{ll} \Gamma^0_{11} = a\dot{a}/(1-\kappa r^2) & \Gamma^1_{11} = \kappa r/(1-\kappa r^2) \\ \Gamma^0_{22} = a\dot{a}r^2 & \Gamma^0_{33} = a\dot{a}r^2\sin^2\theta \\ \Gamma^1_{01} = \Gamma^2_{02} = \Gamma^3_{03} = \dot{a}/a & \\ \Gamma^1_{22} = -r(1-\kappa r^2) & \Gamma^1_{33} = -r(1-\kappa r^2)\sin^2\theta \\ \Gamma^2_{12} = \Gamma^3_{13} = 1/r & \\ \Gamma^2_{33} = -\sin\theta\cos\theta & \Gamma^3_{23} = \cot\theta \ , \end{array} \tag{1.5.36}$$

or related to these by symmetry. Non-zero components of the Ricci tensor are:

$$
\begin{aligned}
R_{00} &= -3\frac{\ddot{a}}{a} && (1.5.37) \\
R_{11} &= \frac{a\ddot{a} + 2\dot{a}^2 + 2\kappa}{1 - \kappa r^2} && (1.5.38) \\
R_{22} &= r^2(a\ddot{a} + 2\dot{a}^2 + 2\kappa) && (1.5.39) \\
R_{33} &= r^2(a\ddot{a} + 2\dot{a}^2 + 2\kappa)\sin^2\theta \ , && (1.5.40)
\end{aligned}
$$

and the Ricci scalar is

$$R = -6\left[\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{\kappa}{a^2}\right] \ . \tag{1.5.41}$$

`EXERCISE:` Verify the components of the Christoffel symbols, the Ricci tensor, and the Ricci scalar given above.

---

4. General Friedmann equations

The first Friedmann equation (1.4.21) becomes:

$$\boxed{\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{\kappa}{a^2}} \ , \tag{1.5.42}$$

and the second Friedmann equation does not change due to κ (`EXERCISE: CHECK`), and we repeat it here for completeness:

$$\boxed{\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P)} \ . \tag{1.5.43}$$

Notice that, in an expanding universe (*i.e.* $\dot{a} > 0$ at all times) filled with ordinary matter (*i.e.* matter satisfying the strong energy condition: $\rho + 3P \geq 0$), Eq. (1.5.43) implies $\ddot{a} < 0$ at all times. This indicates the existence of a singularity in the finite past: $a(t = 0) = 0$. Of course, this conclusion relies on the assumption that general relativity and the Friedmann equations are applicable up to arbitrarily high energies. This assumption is almost certainly not true and it is expected that a quantum theory of gravity will resolve the initial big bang singularity.

B. Dynamics of the FRW universe

1. Terminology

The expansion rate of the FRW universe is characterized by the *Hubble parameter*,

$$\boxed{H(t) = \frac{\dot{a}}{a}} \,. \tag{1.5.44}$$

The expansion rate at the present epoch, $H(t_0)$, is called the *Hubble constant*, H_0[1]. Often you will see the dimensionless number h, where

$$H_0 = 100h \text{ km s}^{-1}\text{Mpc}^{-1} \,. \tag{1.5.45}$$

The astronomical length scale of a megaparsec (Mpc) is equal to 3.0856×10^{24} cm, and h should not be confused with Planck's constant. Observationally, $h \sim 0.7$. Typical cosmological scales are set by the *Hubble length*,

$$d_H = H_0^{-1} c = 9.25 \times 10^{27} h^{-1} \text{ cm} = 3.00 \times 10^3 h^{-1} \text{ Mpc} \,. \tag{1.5.46}$$

The *Hubble time* is,

$$t_H = H_0^{-1} = 3.09 \times 10^{17} h^{-1} \text{ sec} = 9.78 \times 10^9 h^{-1} \text{ yr.} \tag{1.5.47}$$

Since we usually set $c = 1$, H_0^{-1} is referred to as both the Hubble length and the Hubble time. The *deceleration parameter*,

$$q = -\frac{a\ddot{a}}{\dot{a}^2} \,, \tag{1.5.48}$$

measures the rate of change of the expansion.

The density parameter, which counts the energy density from all forms of constituents of the universe, is defined as

$$\boxed{\Omega = \frac{8\pi G}{3H^2}\rho = \frac{\rho}{\rho_{\text{crit}}}} \,, \tag{1.5.49}$$

where the critical density

$$\rho_{\text{crit}} = \frac{3H^2}{8\pi G} \,, \tag{1.5.50}$$

changes with time, and is so called because the Friedmann equation (1.5.42) can be written:

$$\Omega(a) - 1 = \frac{\kappa}{H^2 a^2} \,. \tag{1.5.51}$$

Thus, sign(κ) is defined by sign($\Omega - 1$):

$\rho < \rho_{\text{crit}} \leftrightarrow \Omega < 1 \leftrightarrow \kappa < 0 \leftrightarrow$ open
$\boxed{\rho = \rho_{\text{crit}} \leftrightarrow \Omega = 1 \leftrightarrow \kappa = 0 \leftrightarrow \text{flat}}$
$\rho > \rho_{\text{crit}} \leftrightarrow \Omega > 1 \leftrightarrow \kappa > 0 \leftrightarrow$ closed

The density parameter thus tells us which of the three FRW geometries describes our universe. Our universe is observationally indistinguishable from the flat case. We can further streamline our expressions by treating the contribution of the spatial curvature as a fictitious energy density,

$$\rho_\kappa = -\frac{3\kappa}{8\pi G a^2} \,, \tag{1.5.52}$$

with a corresponding density parameter,

$$\Omega_\kappa = -\frac{\kappa}{H^2 a^2} \,. \tag{1.5.53}$$

---

[1] The "0" subscript is used to denote the present epoch: $t = t_0$, $a(t_0) = a_0 = 1$.

An immediate consequence of the two Friedmann equations is the *continuity equation*, which we previously derived in Eq. (1.4.15) from considering the conservation of the energy-momentum tensor:

$$\frac{d\rho}{dt} + 3H\,(\rho + P) = 0\ . \tag{1.5.54}$$

More heuristically this also follows from the first law of thermodynamics

$$\begin{aligned} dU &= -PdV \\ \mathrm{d}(\rho a^3) &= -P\mathrm{d}(a^3) \qquad \Rightarrow \qquad \frac{d\ln\rho}{d\ln a} = -3(1+w)\ , \end{aligned} \tag{1.5.55}$$

where, w is the equation of state, reminding ourselves of Eq. (1.4.8). The continuity equation (1.4.15) can be integrated to give

$$\boxed{\rho \propto a^{-3(1+w)}}\ . \tag{1.5.56}$$

Together with the Friedmann equation (1.5.42) this leads to the time evolution of the scale factor,

$$\boxed{a \propto t^{2/3(1+w)} \quad \forall \quad w \neq -1}\ . \tag{1.5.57}$$

In particular, constituents of our universe follow the following scalings:

component	w_i	$\rho(a)$	$a(t)$
non-relativistic matter	0	$\propto a^{-3}$	$\propto t^{2/3}$
radiation/relativistic matter	$\frac{1}{3}$	$\propto a^{-4}$	$\propto t^{1/2}$
curvature	$-\frac{1}{3}$	$\propto a^{-2}$	$\propto t$
cosmological constant	-1	$\propto a^{0}$	$\propto \exp(Ht)$

For each species i we define the *present* ratio of the energy density relative to the critical density,

$$\Omega_{i,0} \equiv \frac{\rho_0^i}{\rho_{\mathrm{crit},0}}\ , \tag{1.5.58}$$

and the corresponding equations of state

$$w_i \equiv \frac{P_i}{\rho_i}\ . \tag{1.5.59}$$

This allows one to rewrite the first Friedmann equation (1.5.42) as

$$\left(\frac{H}{H_0}\right)^2 = \sum_i \Omega_{i,0} a^{-3(1+w_i)} + \Omega_{\kappa,0} a^{-2}\ , \tag{1.5.60}$$

which implies the following consistency relation

$$\sum_i \Omega_i + \Omega_\kappa = 1\ . \tag{1.5.61}$$

The second Friedmann equation (1.5.43) evaluated at $t = t_0$ becomes

$$\left(\frac{\ddot{a}}{a}\right)_{t=t_0} = -\frac{H_0^2}{2}\sum_i \Omega_i(1+3w_i)\ . \tag{1.5.62}$$

Observations of the cosmic microwave background (CMB) and the large-scale structure (LSS) find that the universe is flat ($\Omega_\kappa \sim 0$) and composed of 4% atoms, 23%, cold dark matter and 73% dark energy: $\Omega_{\mathrm{b},0} = 0.04$, $\Omega_{\mathrm{cdm},0} = 0.23$, $\Omega_{\Lambda,0} = 0.73$, with $w_\Lambda \approx -1$. In the following, we will sometimes drop the suffix "0" that denotes present-day values of the cosmological parameters unless this is not clear from the context.

C. Matter-radiation equality

The epoch of matter-radiation equality, when $\rho_{\rm r} = \rho_{\rm m}$, has special significance for the generation of large scale structure and the development of CMB anisotropies because perturbations grow at different rates in the two different eras. It is given by,

$$\frac{\rho_{\rm r}}{\rho_{\rm crit}} = \frac{4.15 \times 10^{-5}}{h^2 a^4} \equiv \frac{\Omega_{\rm r}}{a^4} = \frac{\Omega_{\rm m}}{a^3} \ , \tag{1.5.63}$$

where $\Omega_{\rm r}$ and $\Omega_{\rm m}$ are specified at the present epoch, yielding

$$a_{\rm EQ} = \frac{4.15 \times 10^{-5}}{\Omega_{\rm m} h^2} \ . \tag{1.5.64}$$

In terms of redshift,

$$1 + z_{\rm EQ} = 2.4 \times 10^4 \Omega_{\rm m} h^2 \ . \tag{1.5.65}$$

As $\Omega_{\rm m} h^2$ increases, equality is pushed back to higher redshifts and earlier times. It is very important that $z_{\rm EQ}$ is at least a factor of a few larger than the redshift where photons decouple from matter, $z_\star \simeq 1100$, and that the photons decouple when the universe is well into the MD era.

D. Cosmological constant

A characteristic feature of GR is that the source for the gravitational field is the entire energy-momentum tensor. In the absence of gravity, only *changes* in energy from one state to another are measurable; the normalization of the energy is arbitrary. However, in gravitation, the *normalization* of the energy matters. This opens up the possibility of *vacuum energy*: the energy density of empty space. We want the vacuum energy not to pick out a preferred direction. This implies that the associated energy-momentum tensor is Lorentz-invariant in locally inertial coordinates:

$$T_{\mu\nu}^{\rm (vac)} = \rho_{\rm vac} \eta_{\mu\nu} \ . \tag{1.5.66}$$

Generalizing to an arbitrary frame,

$$T_{\mu\nu}^{\rm (vac)} = \rho_{\rm vac} g_{\mu\nu} \ . \tag{1.5.67}$$

Comparing to the perfect fluid energy-momentum tensor, $T_{\mu\nu} = (\rho + P) U_\mu U_\nu - P g_{\mu\nu}$, the vacuum looks like a perfect fluid with an isotropic pressure opposite in sign to the density:

$$P_{\rm vac} = -\rho_{\rm vac} \ . \tag{1.5.68}$$

If we decompose the energy-momentum tensor into a matter piece $T_{\mu\nu}^{\rm (M)}$ plus a vacuum piece (1.5.67), the Einstein equation is:

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi G \left[T_{\mu\nu}^{\rm (M)} + \rho_{\rm vac} g_{\mu\nu} \right] \ . \tag{1.5.69}$$

Einstein tried to get a static universe by adding a *cosmological constant*,

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R - \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}^{\rm (M)} \ , \tag{1.5.70}$$

and this concept is interchangeable with vacuum energy:

$$\rho_\Lambda = \rho_{\rm vac} = \frac{\Lambda}{8\pi G} \ . \tag{1.5.71}$$

Λ has dimensions of [length]$^{-2}$ while $\rho_{\rm vac}$ has units of [energy/volume]. So Λ defines a *scale* (whereas GR is otherwise scale-free). What should be the value of $\rho_{\rm vac}$? There is no known way to precisely calculate this at present, but since the reduced Planck mass is

$$M_P = \frac{1}{\sqrt{8\pi G}} \sim 10^{18} \ {\rm GeV} \ , \tag{1.5.72}$$

one might guess that

$$\rho_{\rm vac}^{(\rm guess)} \sim M_P^4 \sim (10^{18}\ {\rm GeV})^4 \ . \tag{1.5.73}$$

However, this is a dramatically bad guess compared to the observational measurement:

$$\rho_{\rm vac}^{(\rm obs)} \sim (10^{-3}\ {\rm eV})^4 \ , \tag{1.5.74}$$

which is 30 orders of magnitude smaller.

`EXERCISE:` Taking the universe to be at the critical density, $\rho_{\rm crit}$ (*i.e.*, $\Omega = 1$), and $\Omega_\Lambda \sim 0.75$, compute $\rho_{\rm vac}$.

There are three conundrums associated with Λ:

- Why is Λ so small?
- Why is $\Lambda \neq 0$?
- Why is $\Omega_\Lambda \sim \Omega_{\rm m}$?

The last is the so-called *coincidence problem*. Note that

$$\frac{\Omega_\Lambda}{\Omega_{\rm m}} = \frac{\rho_\Lambda}{\rho_{\rm m}} \propto a^3 \ . \tag{1.5.75}$$

The vacuum density rapidly overwhelms the matter density! It is a big coincidence to find ourselves at the epoch where we can observe the transition.

E. Spatial curvature and destiny

In the current universe, the radiation density is significantly lower than the matter density, but both the vacuum and matter are dynamically important. Parameterized as

$$\Omega_\kappa = 1 - \Omega_{\rm m} - \Omega_\Lambda \ , \tag{1.5.76}$$

these densities evolve differently as a function of time:

$$\Omega_\Lambda \propto \Omega_\kappa a^2 \propto \Omega_{\rm m} a^3 \ . \tag{1.5.77}$$

As $a \to 0$ in the past, curvature and vacuum will be negligible and the universe will behave as Einstein-de Sitter till the radiation becomes important. As $a \to \infty$ in the future, curvature and matter will be negligible and the universe will asymptote to de Sitter, unless the scale factor never reaches ∞ because the universe begins to recollapse at a finite time. Some possibilities the evolution of such a universe are:

- $\Omega_\Lambda < 0$: always decelerates and recollapses (as vacuum energy is always going to dominate).
- $\Omega_\Lambda \geq 0$: recollapse is possible if $\Omega_{\rm m}$ is sufficiently large that it halts the universal expansion before Ω_Λ dominates.

To determine the dividing line between perpetual expansion and eventual recollapse, note that collapse requires H to pass through 0 as it changes from positive to negative:

$$H^2 = 0 = \frac{8\pi G}{3}\left(\rho_{{\rm m},0} a_\star^{-3} + \rho_{\Lambda,0} + \rho_{\kappa,0} a_\star^{-2}\right) \ , \tag{1.5.78}$$

where $a_\star$ is the scale-factor at turnaround. Dividing by H_0, using $\Omega_{\kappa,0} = 1 - \Omega_{{\rm m},0} - \Omega_{\Lambda,0}$, and rearranging, we obtain

$$\Omega_{\Lambda,0} a_\star^3 + (1 - \Omega_{{\rm m},0} - \Omega_{\Lambda,0}) a_\star + \Omega_{{\rm m},0} = 0 \ . \tag{1.5.79}$$

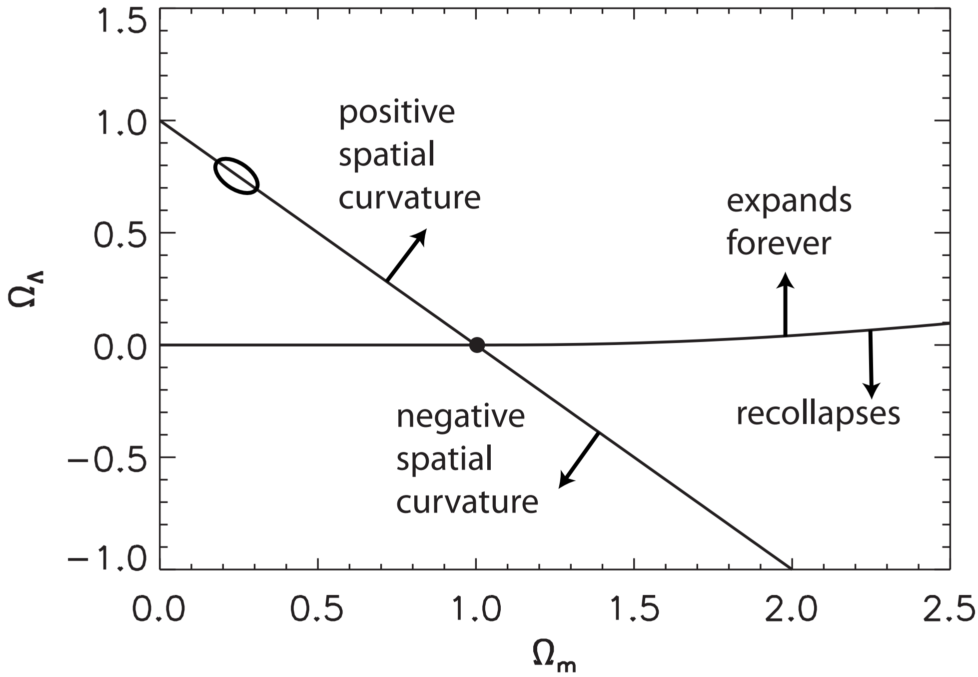

FIG. 8 Properties of universes dominated by matter and vacuum energy, as a function of the density parameters $\Omega_{\rm m}$ and Ω_Λ. The ellipse in the upper-left corner corresponds roughly to the observationally favoured region from a current data compilation as of 2008.

But what we really care about is not really $a_\star$ but the range of $\Omega_{\Lambda,0}$ given $\Omega_{\rm m,0}$ for which there is a real solution to (1.5.79). The range of $\Omega_{\Lambda,0}$ for which the universe will *expand forever* is given by:

$$\Omega_{\Lambda,0} \geq \begin{cases} 0 & 0 \leq \Omega_{\rm m,0} \leq 1 \\ 4\Omega_{\rm m,0}\cos^3\left[\frac{1}{3}\cos^{-1}\left(\frac{1-\Omega_{\rm m,0}}{\Omega_{\rm m,0}}\right)+\frac{4\pi}{3}\right] & \Omega_{\rm m,0} > 1 \end{cases} . \tag{1.5.80}$$

`EXERCISE:` Verify Eq. (1.5.80).

When $\Omega_{\Lambda,0} = 0$,

- open and flat universes ($\Omega_0 = \Omega_{\rm m,0} \leq 1$) will expand forever.
- closed universes ($\Omega_0 = \Omega_{\rm m,0} > 1$) will recollapse.

There is a "folk wisdom" that this correspondence is always true, but it is only true in the *absence of vacuum energy*. The current cosmological data favours $\Omega_{\rm m,0} \sim 0.25, \Omega_{\Lambda,0} \sim 0.75, \Omega_\kappa \sim 0$, which is well into the regime of perpetual expansion, under the assumption that vacuum energy remains truly constant (which it might not). These considerations are illustrated in Fig. 8.

We will end this discussion by noting the difficulty of finding static solutions to the Friedmann equations. To be static, we must have not only $\dot{a} = 0$, but also $\ddot{a} = 0$.

`EXERCISE:` Verify that one can only get a static solution, $\ddot{a} = \dot{a} = 0$, if

$$P = -\frac{\rho}{3} \ , \tag{1.5.81}$$

and the spatial curvature is non-vanishing:

$$\frac{\kappa}{a^2} = \frac{8\pi G}{3}\rho \ . \tag{1.5.82}$$

Because the energy density and pressure must be of the opposite sign, these conditions can't be fulfilled in a universe containing only radiation and matter. Einstein looked for a static solution because at the time, the expansion of the universe had not yet been discovered. He added the cosmological term, whereby one can satisfy the static conditions with

$$\rho_\Lambda = \frac{1}{2}\rho_{\rm m} \quad \text{along with the appropriate positive spatial curvature.} \tag{1.5.83}$$

This *Einstein-static universe* is empirically of little interest today, but extremely useful to theorists, providing the basis for the construction of conformal diagrams.

VI. TIME AND DISTANCE

Measuring distances in an expanding universe is a tricky business! Which distance should one consider? Some obvious definitions immediately come to mind:

- comoving distance (remains fixed as the universe expands).
- physical distance (grows simply because of the expansion).

Frequently, neither of these is what we want; *e.g.* a photon leaves a quasar at $z \sim 6$ when the scale factor was $\frac{1}{7}$ of its present value, and arrives on the earth today, when the universe has expanded by a factor of 7 (see Fig. 9). How can we relate the luminosity of the quasar to the flux we see?

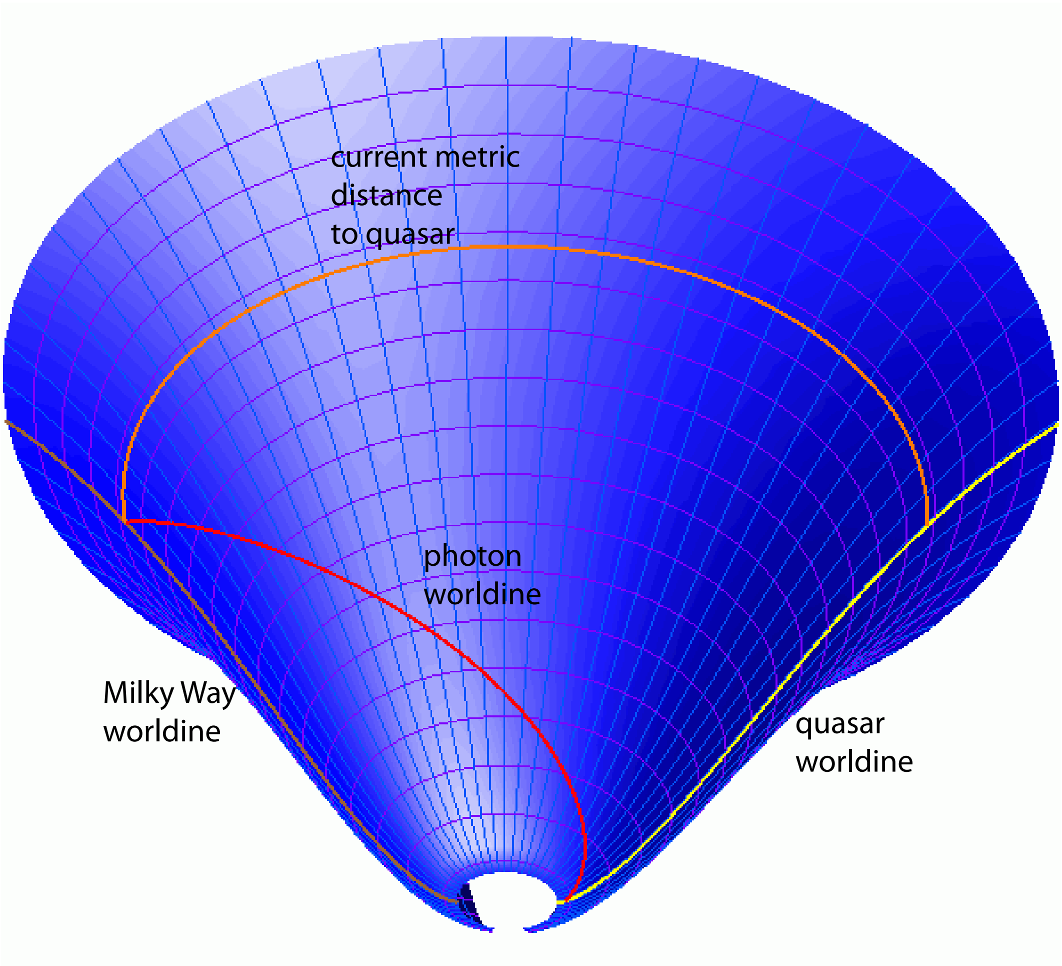

FIG. 9 Euclidean embedding of a part of the ΛCDM spacetime geometry, showing the Milky Way (brown), a quasar at redshift z = 6.4 (yellow), light from the quasar reaching the Earth after approximately 12 billion years (red), and the present-era metric distance to the quasar of approximately 28 billion light years (orange). Lines of latitude (purple) are lines of constant cosmological time, spaced by 1 billion years; lines of longitude (cyan) are worldlines of objects moving with the Hubble flow, spaced by 1 billion light years in the present era (less in the past and more in the future). Figure credit: Ben Rudiak-Gould / Wikimedia Commons.

A. Conformal time

The fundamental measure from which all others may be calculated is the *distance on the comoving grid*. If the universe is flat, as we will assume throughout most of these lectures, computing distances on the comoving grid is easy.

One very important comoving distance is the distance travelled by light since $t = 0$ (in the absence of interactions). Recalling that we are working in units with $c = 1$, in time dt, light travels a distance $dx = \frac{dt}{a}$; thus, the total comoving distance light travels is:

$$\eta \equiv \int_0^t \frac{dt'}{a(t')} \; . \tag{1.6.1}$$

No information could have propagated faster than η on the comoving grid since the beginning of time; thus η is called the *causal horizon* or *comoving horizon*. A related concept is the *particle horizon* d_H, the proper radius travelled by light since $t = 0$:

$$d_H \equiv a(t) \int_0^t \frac{dt'}{a(t')} = a(\eta)\eta \; . \tag{1.6.2}$$

Regions separated by distances $> d_H$ are not causally connected; if they appear similar, we should be suspicious! (*cf.* the cosmic microwave background – see later!). We can think of η (which increases monotonically) as a time variable and call it the *conformal time*. In terms of η, the FRW metric becomes

$$\boxed{ds^2 = a^2(\eta)\left[d\eta^2 - \frac{dr^2}{1-\kappa r^2} - r^2 d\Omega^2\right]} \; . \tag{1.6.3}$$

Just like $\{t, T, z, a\}$, η can be used to discuss the evolution of the universe. η is the most useful time variable for most purposes! In the analysis of the evolution of perturbations, we will use it instead of t. Conformal spacetime diagrams are easy to construct in terms of η. Consider radial null geodesics in a flat FRW spacetime:

$$ds^2 = a^2(d\eta^2 - dr^2) \equiv 0 \quad \Longrightarrow \quad d\eta = \pm\frac{dt}{a(t)} \equiv \pm dr \; . \tag{1.6.4}$$

In conformal coordinates, null geodesics (photon worldlines) are always at 45° angles, and light cones are Minkowskian since the metric is *conformally flat*: $g_{\mu\nu} = a^2\eta_{\mu\nu}$.

In simple cases, η can be expressed analytically in terms of a. In particular, during radiation domination (RD) and matter domination (MD),

$$\text{RD}: \qquad \rho \propto a^{-4}, \; \eta \propto a \tag{1.6.5}$$

$$\text{MD}: \qquad \rho \propto a^{-3}, \; \eta \propto \sqrt{a} \; . \tag{1.6.6}$$

`EXERCISE:` Show that the conformal time as a function of scale-factor in a flat universe containing only matter and radiation is

$$\frac{\eta}{\eta_0} = \sqrt{a + a_{\rm EQ}} - \sqrt{a_{\rm EQ}} \; . \tag{1.6.7}$$

where $a_{\rm EQ}$ denotes the epoch of matter-radiation equality.

B. Lookback distance and lookback time

Another important comoving distance is the distance between us and a distant emitter, the *lookback distance*. The comoving distance to an object at scale factor a (or redshift $z = \frac{1}{a} - 1$) is:

$$d_{\rm lookback}(a) = \int_{t(a)}^{t_0} \frac{dt'}{a(t')} = \int_a^1 \frac{da'}{a^2(t')H(a')} \; , \tag{1.6.8}$$

where we have used $\frac{da}{dt} = aH$. Typically, we can see objects out to $z \lesssim 6$. At these late times, radiation can be ignored. During matter domination, $H \propto a^{-3/2}$, so

$$\text{FLAT MD}: \quad d_{\rm lookback}(a) = \frac{2}{H_0}\left[1 - \sqrt{a}\right] \tag{1.6.9}$$

$$d_{\rm lookback}(z) = \frac{2}{H_0}\left[1 - \frac{1}{\sqrt{1+z}}\right] \; . \tag{1.6.10}$$

For small z, $d_{\rm lookback} \to \frac{z}{H_0}$, which is the *Hubble Law*. At very early times, $z \gg 1$, we find the limit $d_{\rm lookback} \to \frac{2}{H_0}$.
Similarly, one can define the *lookback time*, elapsed between now and when light from redshift z was emitted:

$$t_{\rm lookback}(a) = \int_{t(a)}^{t_0} dt' = \int_a^1 \frac{da'}{a(t')H(a')} \ . \tag{1.6.11}$$

For a flat, matter-dominated universe, the lookback time to redshift z is:

$$\text{FLAT MD}: \quad t_{\rm lookback}(z) = \frac{2}{3H_0}\left[1-(1+z)^{-3/2}\right] \ . \tag{1.6.12}$$

The total age of a matter-dominated universe is obtained by letting $z \to \infty$:

$$t_0(\text{FLAT MD}) = \frac{2}{3H_0} \ . \tag{1.6.13}$$

For universes that are not totally matter-dominated, the factor of $\frac{2}{3}$ will not be quite right, but for reasonable values of the cosmological parameters, we usually get $t_0 \sim H_0^{-1}$.

C. Instantaneous physical distance

Another distance we might want to know is the distance between us and the location of a distant object along our current spatial hypersurface. Let us write the FRW metric in the form

$$ds^2 = dt^2 - a^2(t)R_0^2\left[d\chi^2 + S_k^2(\chi)d\Omega^2\right] \ , \tag{1.6.14}$$

where $S_k(\chi)$ is defined by (1.5.27) and $k \in \{+1, 0, -1\}$. In this form, the instantaneous physical distance d_P as measured at time t between us at $\chi = 0$ and an object at comoving radial coordinate χ is,

$$d_P(t) = a(t)R_0\chi \ , \tag{1.6.15}$$

where χ remains constant because we assume that both we and the observed object are perfectly comoving (they might not be, but it is trivial to include the corrections due to so-called "peculiar velocities"). As expected, this definition of distance also leads to the Hubble Law when the redshift is small,

$$v = \dot{d}_P = \dot{a}R_0\chi = \frac{\dot{a}}{a}d_P \quad \longrightarrow \quad v = H_0 d_P \tag{1.6.16}$$

when evaluated today.
The instantaneous physical distance, while a convenient construct, is not that useful, because observations always refer to events on our past lightcone, not on our current spatial hypersurface. For various kinds of observations, we can define a kind of distance that is what we *would* infer if space were Euclidean and the universe were not expanding, and relate it to observables in the FRW universe.

D. Luminosity distance

A classic way of measuring distances in astronomy is to measure the flux from an object of known luminosity, a *standard candle*. Let us neglect expansion for a moment, and consider the observed flux F at a distance d_L from a source of known luminosity L:

$$F = \frac{L}{4\pi d_L^2} \ . \tag{1.6.17}$$

This definition comes from the fact that in flat space, for a source at distance d, the flux over the luminosity is just the inverse of the area of a sphere centred around the source, $1/A(d) = 1/4\pi d^2$. In an FRW universe, however, the flux will be diluted. Conservation of photons tells us that all of the photons emitted by the source will eventually pass through a sphere at a comoving distance χ from the emitter. But the flux is diluted by two additional effects: the individual photons redshift by a factor $(1+z)$, and the photons hit the sphere less frequently, since two photons emitted a time δt apart will be measured at a time $(1+z)\delta t$ apart. Therefore we will have

$$\frac{F}{L} = \frac{1}{(1+z)^2 A} \ . \tag{1.6.18}$$

The area A of a sphere centred at a comoving distance χ can be derived from the coefficient of $d\Omega^2$ in (1.6.14), yielding

$$A = 4\pi R_0^2 S_k^2(\chi) \ , \tag{1.6.19}$$

where we have set $a(t) = 1$ because the photons are being observed today. Comparing with (1.6.17), we obtain the *luminosity distance*:

$$d_L = (1+z) R_0 S_k(\chi) \ . \tag{1.6.20}$$

Here, we must point out a caveat: the observed luminosity is related to emitted luminosity at a *different wavelength*. Here, we have assumed that the detector counts all photons.

The luminosity distance d_L is something we might hope to measure, since there are some astrophysical sources which are standard candles. But χ is not an observable, so we should rephrase it in terms of something we can measure. On a radial null geodesic, we have

$$0 = ds^2 = dt^2 - a^2 R_0^2 d\chi^2 \ , \tag{1.6.21}$$

or

$$\chi = \frac{1}{R_0}\int \frac{dt}{a} = \frac{1}{R_0}\int \frac{da}{a^2 H(a)} \ . \tag{1.6.22}$$

It's conventional to convert the scale factor to redshift using $a = 1/(1+z)$, so we have

$$\chi(z) = \frac{1}{R_0}\int_0^z \frac{dz'}{H(z')} \ , \tag{1.6.23}$$

leading to the luminosity distance,

$$d_L = (1+z) R_0 S_k \left[\frac{1}{R_0}\int_0^z \frac{dz'}{H(z')}\right] \ . \tag{1.6.24}$$

Note that R_0 drops out when $k = 0$, which is good because in that case it is a completely arbitrary parameter. Even when it is not arbitrary, it is still more common to speak in terms of $\Omega_{\kappa,0} = -k/R_0^2 H_0^2$, which can either be determined though measurements of the spatial curvature, or by measuring the matter density and using $\Omega_{\kappa,0} = 1 - \Omega_{\mathrm{m},0}$. In terms of this parameter, we have

$$R_0 = H_0^{-1}\sqrt{-k/\Omega_{\kappa,0}} = \frac{H_0^{-1}}{\sqrt{|\Omega_{\kappa,0}|}} \ . \tag{1.6.25}$$

Thus we can write the luminosity distance in terms of measurable cosmological parameters as

$$\boxed{d_L = (1+z)\frac{H_0^{-1}}{\sqrt{|\Omega_{\kappa,0}|}} S_k\left[H_0\sqrt{|\Omega_{\kappa,0}|}\int_0^z \frac{dz'}{H(z')}\right]} \ , \tag{1.6.26}$$

where the integral can be evaluated by making use of the Friedmann equation. Though it appears unwieldy, this equation is of fundamental importance in cosmology. Given the observables H_0 and $\Omega_{i,0}$, we can calculate d_L to an object any redshift z; conversely, we can measure $d_L(z)$ for objects at a range of redshifts, and from that extract H_0 or the $\Omega_{i,0}$.

E. Angular diameter distance

Another classic distance measurement in astronomy is to measure the angle $\delta\theta$ subtended by an object of known physical size ℓ, known as a *standard ruler*. The *angular diameter distance* is then defined as,

$$d_A = \frac{\ell}{\delta\theta} \ , \tag{1.6.27}$$

where $\delta\theta$ is small. At the time when the light was emitted, when the universe had scale factor a, the object was at redshift z at comoving coordinate χ (assuming again that we are at $\chi = 0$). Hence, from the angular part of the metric, $\ell = a R_0 S_k(\chi)\delta\theta$, and comparing with (1.6.27) we have the *angular diameter distance*

$$d_A = \frac{R_0 S_k(\chi)}{1+z} \ . \tag{1.6.28}$$

Fortunately, the unwieldy dependence on cosmological parameters is common to all distance measures, and we are left with a simple dependence on redshift:

$$\boxed{d_L = (1+z)^2 d_A}\ . \tag{1.6.29}$$

Note that d_A is equal to the comoving distance at low redshift! But it actually decreases at very large redshift. In a flat universe, objects at large redshift appear *larger* than they would at intermediate redshift.

Both distances: d_A, d_L are *larger* in a universe with a cosmological constant than in one without. This follows since the energy density, and hence the expansion rate, is smaller in a Λ universe. The universe was therefore expanding more slowly early on, and light had more time to travel from distant objects to us. Distant objects will therefore appear *fainter* in a Λ-dominated universe than if the universe was MD today. This observation (using Type Ia supernovae as standardizable candles) is exactly what lead to the discovery of dark energy in the 1990's.

VII. PARTICLES AND FIELDS IN COSMOLOGY

A. General particle motion

We previously considered the redshifting of a massless particle. Let us consider the case of a general particle for completeness. This is, of course, governed by the geodesic equation

$$\frac{du^\mu}{ds} + \Gamma^\mu_{\nu\sigma} u^\nu u^\sigma = 0 \; , \tag{1.7.1}$$

where $u^\mu = \frac{dx^\mu}{ds}$. In the metric (1.5.35), we have

$$u^\mu u_\mu = g_{00}(u^0)^2 + g_{ij} u^i u^j \equiv (u^0)^2 - |\tilde{u}|^2 = \begin{cases} 0 & \text{photon} \\ 1 & \text{massive particle} \end{cases} \tag{1.7.2}$$

where $|\tilde{u}|^2 = -g_{ij}u^i u^j$ is the physical velocity[2]. The latter result comes from the fact that for the 4-momentum $p^\mu = (E, \vec{p})$, $p^\mu p_\mu = 0$ for photons, and the on-shell condition $p^\mu p_\mu = m^2$ holds for massive particles, with $p^\mu = m u^\mu$. For both cases,

$$u^0 \frac{du^0}{ds} = |\tilde{u}| \frac{d|\tilde{u}|}{ds} \; . \tag{1.7.3}$$

For the metric (1.5.35), the relevant Christoffel symbol can be expressed as $\Gamma^0_{ij} = -\frac{\dot{a}}{a} g_{ij}$, so the 0th component of (1.7.1) becomes

$$\frac{du^0}{ds} - \frac{\dot{a}}{a} g_{ij} u^i u^j = \frac{du^0}{ds} + \frac{\dot{a}}{a}|\tilde{u}|^2 = 0 \; , \tag{1.7.4}$$

and thus by (1.7.3),

$$\frac{|\tilde{u}|}{u^0}\frac{d|\tilde{u}|}{ds} + \frac{\dot{a}}{a}|\tilde{u}|^2 = 0 \; . \tag{1.7.5}$$

But $u^0 = \frac{dt}{ds}$, so $|\dot{\tilde{u}}| + \frac{\dot{a}}{a}|\tilde{u}| = 0$, or

$$\frac{|\dot{\tilde{u}}|}{|\tilde{u}|} = -\frac{\dot{a}}{a} \; . \tag{1.7.6}$$

Hence, for both massless *and* massive particles, momentum always redshifts as

$$\boxed{|\tilde{u}| \propto \frac{1}{a}} \; . \tag{1.7.7}$$

The particle therefore slows down with respect to the comoving coordinates as the universe expands. In fact this is an actual slowing down, in the sense that a gas of particles with initially high relative velocities will cool down as the universe expands.

B. Classical field theory

When we make the transition from SR to GR, the Minkowski metric $\eta_{\mu\nu}$ is promoted to a dynamical tensor field, $g_{\mu\nu}(x)$. GR is an example of a *classical field theory.* Let's get a feel for how such theories work by considering classical fields in a flat spacetime. We will not discuss *quantum fields* here, though this will become relevant in the third part of the course when the origin of primordial perturbations is discussed.

---

[2] Recall again that $\vec{p}$ is the comoving momentum, and the physical momentum $\tilde{p}$ measuring changes in physical distance is given by $|\tilde{p}| = \sqrt{|g_{ij}p^i p^j|} = a|\vec{p}|$. Similarly, for 4-velocity $u^\mu = (1, \vec{u})$, we relate the comoving velocity $\vec{u}$ to the physical velocity as $|\tilde{u}| = \sqrt{|g_{ij}u^i u^j|} = a|\vec{u}|$.

Begin with the familiar example of the classical mechanics of a single particle in 1D with coordinate $q(t)$. Equations of motion for such a particle comes from using the *principle of least action*: search for critical points (as a function of the trajectory) of an *action* S,

$$S = \int dt L(q, \dot{q}) \ , \tag{1.7.8}$$

where $L(q, \dot{q})$ is the Lagrangian. The Lagrangian in point-particle interactions is typically of the form $L = K - V$, where K is the kinetic energy and V is the potential energy. Using the calculation of variations procedure (*cf.* any advanced classical mechanics textbook), the critical points of the action, *i.e.* trajectories $q(t)$ for which the action S remains stationary under small variations, are those that satisfy the *Euler-Lagrange equations*:

$$\boxed{\frac{\partial L}{\partial q} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) = 0} \ . \tag{1.7.9}$$

For example, $L = \frac{1}{2}\dot{q}^2 - V(q)$ leads to the equation of motion $\ddot{q} = -\frac{dV}{dq}$. For field theory, we replace the single coordinate $q(t)$ by a set of spacetime-dependent *fields* $\Phi^i(x^\mu)$, and the action becomes a *functional* of these fields. i labels individual fields. A functional is a function of an infinite number of variables, *e.g.* the values of a field in some region of spacetime.

In field theory, the Lagrangian can be expressed as an integral over the space of a *Lagrangian density*, $\mathcal{L}$, which is a function of the fields Φ^i and their spacetime derivatives $\partial_\mu \Phi^i$:

$$L = \int d^3x \, \mathcal{L}(\Phi^i, \partial_\mu \Phi^i) \ . \tag{1.7.10}$$

Then the action becomes,

$$S = \int dt L = \int d^4x \, \mathcal{L}(\Phi^i, \partial_\mu \Phi^i) \ . \tag{1.7.11}$$

$\mathcal{L}$ is a Lorentz scalar. It is most convenient to define a field theory by specifying the Lagrange density, from which all equations of motion can be derived. The Euler-Lagrange equations again come from requiring that S be invariant under small variations of the field,

$$\Phi^i \to \Phi^i + \delta\Phi^i, \quad \partial_\mu \Phi^i \to \partial_\mu \Phi^i + \delta(\partial_\mu \Phi^i) = \partial_\mu \Phi^i + \partial_\mu(\delta\Phi^i) \ . \tag{1.7.12}$$

The expression for variation in $\partial_\mu \Phi^i$ is simply the derivative of the variation of Φ^i. Since $\delta\Phi^i$ is assumed to be small, we can Taylor-expand the Lagrangian under this variation,

$$\begin{aligned} \mathcal{L}(\Phi^i, \partial_\mu \Phi^i) \ &\to \ \mathcal{L}(\Phi^i + \delta\Phi^i, \partial_\mu \Phi^i + \partial_\mu \delta\Phi^i) \ , \\ &= \ \mathcal{L}(\Phi^i, \partial_\mu \Phi^i) + \frac{\partial \mathcal{L}}{\partial \Phi^i}\delta\Phi^i + \frac{\partial \mathcal{L}}{\partial(\partial_\mu \Phi^i)}\partial_\mu(\delta\Phi^i) \ . \end{aligned} \tag{1.7.13}$$

Correspondingly, the action goes to $S \to S + \delta S$, with

$$\delta S = \int d^4x \left[\frac{\partial \mathcal{L}}{\partial \Phi^i}\delta\Phi^i + \frac{\partial \mathcal{L}}{\partial(\partial_\mu \Phi^i)}\partial_\mu(\delta\Phi^i)\right] \ . \tag{1.7.14}$$

EXERCISE: Factor out the $\delta\Phi^i$ term from the integrand, by integrating the second term by parts. You will obtain one term which is a total derivative – the integral of something of the form $\partial_\mu V^\mu$ – that can be converted to a surface term by the four-dimensional version of Stokes' Theorem. Since we are considering variational problems, we can choose to consider variations that vanish at the boundary, along with their derivatives. It is therefore traditional in such contexts to integrate by parts with complete impunity, always ignoring the boundary conditions. Sometimes this is not okay, but fortunately we will not encounter such situations in this course. Assuming that the variations and their derivatives vanish at the boundaries, show that

$$\delta S = \int d^4x \left[\frac{\partial \mathcal{L}}{\partial \Phi^i} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \Phi^i)} \right) \right] \delta\Phi^i \ . \tag{1.7.15}$$

The functional derivative $\delta S/\delta\Phi^i$ of a functional S with respect to a function Φ^i is defined to satisfy

$$\delta S = \int d^4x \frac{\delta S}{\delta \Phi^i} \delta\Phi^i \tag{1.7.16}$$

when such an expression is valid. We can therefore express the notion that S is at a critical point by saying that the functional derivative vanishes. Finally we arrive at the Euler-Lagrange equations of motion for a field theory in flat spacetime:

$$\boxed{\frac{\delta S}{\delta \Phi^i} = \frac{\partial \mathcal{L}}{\partial \Phi^i} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \Phi^i)} \right) = 0} \ . \tag{1.7.17}$$

There are several benefits of introducing the Lagrangian formulation:

- The simplicity of positing a single scalar-valued function of spacetime, the Langrange density, rather than a number of (perhaps tensor-valued) equations of motion.
- Demanding that the action be invariant under a symmetry assures that the dynamics respect the symmetry too.
- The action leads via a direct procedure to a unique energy-momentum tensor.

C. Energy-momentum tensor from the action

As promised earlier, we will now consider the last point. First, we need to generalize the previous discussion to curved space. Recalling our experience with the geodesic equation, first we will replace the partial derivative by the covariant derivative, defined *e.g.*

$$\nabla_\mu T^{\alpha\beta} \equiv \frac{\partial T^{\alpha\beta}}{\partial x^\mu} + \Gamma^\alpha_{\gamma\mu} T^{\gamma\beta} + \Gamma^\beta_{\gamma\mu} T^{\alpha\gamma} \equiv T^{\alpha\beta}{}_{;\mu} \ , \tag{1.7.18}$$

$$\nabla_\mu T_{\alpha\beta} \equiv \frac{\partial T_{\alpha\beta}}{\partial x^\mu} - \Gamma^\gamma_{\alpha\mu} T_{\gamma\beta} - \Gamma^\gamma_{\beta\mu} T_{\alpha\gamma} \equiv T_{\alpha\beta;\mu} \ , \tag{1.7.19}$$

$$\nabla_\mu T^{\alpha} \equiv \frac{\partial T^{\alpha}}{\partial x^\mu} + \Gamma^\alpha_{\gamma\mu} T^{\gamma} \equiv T^{\alpha}{}_{;\mu} \ , \tag{1.7.20}$$

$$\nabla_\mu T_{\alpha} \equiv \frac{\partial T_{\alpha}}{\partial x^\mu} - \Gamma^\gamma_{\alpha\mu} T_{\gamma} \equiv T_{\alpha;\mu} \ , \tag{1.7.21}$$

$$\nabla_\mu T \equiv \frac{\partial T}{\partial x^\mu} \equiv T_{\ ;\mu} \ , \tag{1.7.22}$$

$$g_{\alpha\beta;\mu} = 0 \ . \tag{1.7.23}$$

Now in n dimensions,

$$S = \int d^n x \, \mathcal{L}(\Phi^i, \nabla_\mu \Phi^i) \ . \tag{1.7.24}$$

Note that, since $d^n x$ is a density rather than a tensor, $\mathcal{L}$ is also a density. We typically write,

$$\mathcal{L} = \sqrt{-g}\,\hat{\mathcal{L}}\ , \tag{1.7.25}$$

where $\hat{\mathcal{L}}$ is a scalar, and $g = \det g_{\mu\nu}$. The associated Euler-Lagrange equations make use of the scalar $\hat{\mathcal{L}}$, and are like those in flat space but with covariant instead of partial derivatives:

$$\frac{\partial \hat{\mathcal{L}}}{\partial \Phi} - \nabla_\mu \left(\frac{\partial \hat{\mathcal{L}}}{\partial(\nabla_\mu \Phi)} \right) = 0\ . \tag{1.7.26}$$

e.g., the curved space generalization of the action for a single scalar field is:

$$S_\phi = \int d^n x \sqrt{-g} \left[\frac{1}{2} g^{\mu\nu} (\nabla_\mu \phi)(\nabla_\nu \phi) - V(\phi) \right]\ , \tag{1.7.27}$$

which would lead to an equation of motion,

$$\Box \phi + \frac{dV}{d\phi} = 0\ , \tag{1.7.28}$$

where

$$\Box \phi = \nabla^\nu \nabla_\nu \phi = g^{\mu\nu} \nabla_\mu \nabla_\nu \phi\ . \tag{1.7.29}$$

With that warm-up, let's think about the construction of an action for GR. The dynamical variable is now $g_{\mu\nu}$, the metric. What scalars can we make out of the metric to serve as a Lagrangian? Since we know that the metric can be set equal to its canonical form (*i.e.*, $g_{\mu\nu} = \eta_{\mu\nu}$) and its derivatives set to zero at any one point, any non-trivial scalar must involve at least second derivatives of the metric.

We have already encountered the *Ricci scalar*. It turns out to be the only independent scalar constructed from the metric, which is no higher than second order in its derivatives. Hilbert figured that this was the simplest possible choice for a Lagrangian, and proposed:

$$\boxed{S_H = -\int d^n x \sqrt{-g} R \qquad \text{(Einstein–Hilbert Action)}}\ . \tag{1.7.30}$$

He was right! Beware the sign convention of the metric signature here. If you see this expression with no minus-sign in the literature, remember that $R_{\mu\nu}$ is invariant under sign-change of the metric, and hence $R = g^{\mu\nu} R_{\mu\nu}$ flips sign under sign change of the metric. Cutting a very long story short, varying the action,

$$\delta S_H = -\int d^n x \sqrt{-g} \left[R_{\mu\nu} - \frac{1}{2} R\, g_{\mu\nu} \right] \delta g^{\mu\nu}\ , \tag{1.7.31}$$

we find that at stationary points,

$$-\frac{1}{\sqrt{-g}} \frac{\delta S_H}{\delta g^{\mu\nu}} = R_{\mu\nu} - \frac{1}{2} R\, g_{\mu\nu} = 0\ . \tag{1.7.32}$$

Voila – we have recovered Einstein's equation in a vacuum! We got the result in a vacuum because we only included the gravitational part of the action, not the matter fields. To include them, consider

$$S = \frac{1}{16\pi G} S_H + S_M\ , \tag{1.7.33}$$

where S_M is the matter action, and we have presciently normalized the gravitational action to get the right answer. Again, cutting out the details of the variational procedure, which you are welcome to work through on your own, one obtains,

$$\frac{1}{\sqrt{-g}} \frac{\delta S}{\delta g^{\mu\nu}} = -\frac{1}{16\pi G} \left[R_{\mu\nu} - \frac{1}{2} R\, g_{\mu\nu} \right] + \frac{1}{\sqrt{-g}} \frac{\delta S_M}{\delta g^{\mu\nu}} = 0\ . \tag{1.7.34}$$

Let's boldly define the energy-momentum tensor as:

$$\boxed{T_{\mu\nu} = \frac{2}{\sqrt{-g}}\frac{\delta S_M}{\delta g^{\mu\nu}}} \; . \tag{1.7.35}$$

We immediately see that this definition leads to Einstein's equation:

$$R_{\mu\nu} - \frac{1}{2}R\, g_{\mu\nu} = 8\pi G\, T_{\mu\nu} \; . \tag{1.7.36}$$

Consider again the action (1.7.27) for a scalar field. Now vary this action with respect to $g_{\mu\nu}$, not ϕ:

$$\delta S_\phi = \int d^n x\sqrt{-g}\,\delta g^{\mu\nu}\left[\frac{1}{2}\nabla_\mu\phi\nabla_\nu\phi - \frac{g_{\mu\nu}}{2}\left(\frac{g^{\rho\sigma}}{2}\nabla_\rho\phi\nabla_\sigma\phi - V(\phi)\right)\right] \; , \tag{1.7.37}$$

we obtain from (1.7.35) the energy-momentum tensor for a scalar field,

$$\boxed{T^{(\phi)}_{\mu\nu} = \nabla_\mu\phi\nabla_\nu\phi - g_{\mu\nu}\left[\frac{1}{2}g^{\rho\sigma}\nabla_\rho\phi\nabla_\sigma\phi - V(\phi)\right]} \; . \tag{1.7.38}$$

Here, it is worth pointing out that you will find different sign conventions to (1.7.38) in the cosmology literature. This can be traced to the fact that most of the cosmology literature uses the opposite metric signature to ours: $(-,+,+,+)$. For a homogeneous field ($\partial_i\phi = 0$), we don't need the covariant derivative ("semi-colon goes to a comma rule"), and this reduces to

$$T^{(\phi)}_{\mu\nu} = \partial_\mu\phi\partial_\nu\phi - g_{\mu\nu}\left[\frac{1}{2}g^{\rho\sigma}\partial_\rho\phi\partial_\sigma\phi - V(\phi)\right] \; . \tag{1.7.39}$$

`EXERCISE:` If you want to double check this derivation via the variational procedure, you might find it useful to note that the Kronecker delta $\delta^\mu{}_\nu$ is unchanged under any variation, and first derive the following results:

$$\delta g_{\mu\nu} = -g_{\mu\rho}g_{\nu\sigma}\delta g^{\rho\sigma} \; , \tag{1.7.40}$$

$$\delta g = -g(g_{\mu\nu}\delta g^{\mu\nu}) \; , \tag{1.7.41}$$

where the latter expression requires the variation of the identity $\ln(\det\mathbf{M}) = \mathrm{Tr}(\ln\mathbf{M})$, and $\mathbf{M}$ is a square matrix with non-vanishing determinant, with $\exp(\ln\mathbf{M}) = \mathbf{M}$. It is helpful to remember the cyclic property of the trace. You will need to show also that the term $\propto \delta R_{\mu\nu}$ leads to an integral with respect to the natural volume element of the covariant divergence of a vector. By Stokes' Theorem, this is equal to a boundary condition at infinity, which contributes nothing to the total variation. Consider subtleties which may affect this conclusion.

In Minkowski space there is an alternative definition for the "canonical energy-momentum tensor" (often discussed in books on electromagnetism or field theory). In this context, energy-momentum conservation arises as a consequence of symmetry of the Lagrangian under spacetime translations, via *Noether's theorem.* Under this procedure to a Lagrangian that depends on some fields Φ^i and their first derivatives $\partial_\mu\Phi^i$ (in flat spacetime), one obtains the canonical energy momentum tensor:

$$T^{\mu\nu}_{(\text{canonical})} = \frac{\delta\mathcal{L}}{\delta(\partial_\mu\Phi^i)}\partial^\nu\Phi^i - \eta^{\mu\nu}\mathcal{L} = \partial^\mu\Phi^i\partial^\nu\Phi^i - \eta^{\mu\nu}\mathcal{L} \; , \tag{1.7.42}$$

where a sum over i is implied. We will not discuss this procedure further here, firstly because it does not always generalize to a curved spacetime, and also because (1.7.35) is what appears on the RHS of Einstein's equation when it is derived from an action.

Before we move on, let us remember that we arrived at the Hilbert Lagrangian $\hat{\mathcal{L}}_H = R$ by looking for the simplest possible scalar constructed from the metric. Of course, an even simpler one is a constant, so

$$S = \int d^4x \sqrt{-g} \left[-\frac{1}{16\pi G} \left(R + 2\Lambda \right) + \hat{\mathcal{L}}^{(\mathrm{M})} \right] \ , \tag{1.7.43}$$

giving

$$\hat{\mathcal{L}}^{(\mathrm{vac})} = -\rho_{\mathrm{vac}} \ . \tag{1.7.44}$$

So it is easy to introduce vacuum energy; however, we have no insight into its expected value, since it enters as an arbitrary constant.

VIII. BASICS OF INFLATION

A. Shortcomings of the Standard Big Bang Theory

A number of fundamental questions about the universe are raised by the Big Bang theory:

1. Why is the universe spatially flat on large scales?
2. Why is the universe so homogeneous on large scales?
3. What started the "Big Bang"?
4. What was the origin of the primordial fluctuations which lead to the complex structures observed in the universe today?

Let us discuss these puzzles in more detail.

1. Homogeneity Problem

We previously derived the FRW metric assuming the homogeneity and isotropy of the universe. Why is this a good assumption? This is particularly surprising given that inhomogeneities are gravitationally unstable and therefore grow with time (*cf.* the third part of the course). Observations of the CMB show that the inhomogeneities were much smaller in the past (at decoupling) than today. One thus expects that these inhomogeneities were even smaller at earlier times. How do we explain the homogeneity of the early universe?

2. Flatness Problem

Why is the universe so closely approximated by flat Euclidean space? To understand the severity of the problem in more detail consider the Friedmann equation in the form of (1.5.51), reproduced here for completeness:

$$\Omega(a) - 1 = \frac{\kappa}{(aH)^2} \, . \tag{1.8.1}$$

In the standard Big Bang cosmology containing just matter and radiation, the comoving Hubble radius, $(aH)^{-1}$, grows with time and from (1.5.51), the quantity $|\Omega - 1|$ must thus diverge with time. $\Omega = 1$ is an *unstable fixed point*. Therefore, in standard Big Bang cosmology, the near-flatness observed today ($\Omega_0 \sim 1$) requires an extreme fine-tuning of Ω close to 1 in the early universe. More specifically, one finds that the deviation from flatness at Big Bang nucleosynthesis (BBN, ~ 0.1 MeV), during the GUT era ($\sim 10^{15}$ GeV) and at the Planck scale ($\sim 10^{19}$ GeV), respectively has to satisfy the following conditions

$$|\Omega(a_{\rm BBN}) - 1| \;\le\; \mathcal{O}(10^{-16}) \tag{1.8.2}$$
$$|\Omega(a_{\rm GUT}) - 1| \;\le\; \mathcal{O}(10^{-55}) \tag{1.8.3}$$
$$|\Omega(a_{\rm Pl}) - 1| \;\le\; \mathcal{O}(10^{-61}) \, . \tag{1.8.4}$$

Another way of understanding the flatness problem is by differentiating (1.5.51) and using the continuity equation in the form of (1.5.55) to obtain

$$\frac{d\Omega}{d[\ln a]} = (1+3w)\Omega(\Omega - 1) \, . \tag{1.8.5}$$

This makes it apparent that $\Omega = 1$ is an unstable fixed point if the strong energy condition is satisfied

$$\frac{d|\Omega - 1|}{d[\ln a]} > 0 \quad \Longleftrightarrow \quad 1 + 3w > 0 \, . \tag{1.8.6}$$

Again, why is $\Omega_0 \sim \mathcal{O}(1)$ and not much smaller or much larger?

EXERCISE: $\Omega(t) = \frac{8\pi G\rho(t)}{3H^2(t)}$. Assume $\Omega_0 \sim 0.3$ today, and $\Omega_\Lambda = 0$. Plot $\Omega(t) - 1$ as a function of $a(t)$ assuming that the universe was first radiation-dominated and then matter-dominated. How close is $\Omega(t) - 1$ to zero at the Planck epoch, $a \sim 10^{-32}$? This fine-tuning of the initial conditions is known as the *flatness problem.* If not for fine-tuning, an open universe would be *obviously* open, and Ω_0 would be almost exactly zero today. Instead, we observe that $|\Omega_0 - 1| \lesssim \mathcal{O}(0.01)$ today.

---

3. Horizon Problem

The *comoving Hubble radius*, $(aH)^{-1}$, characterizes the fraction of comoving space in causal contact. During the Big Bang expansion $(aH)^{-1}$ *grows* monotonically and the *fraction of the universe in causal contact increases with time.*

Consider the CMB today. It is very close to isotropic and its temperature is the same everywhere to one part in 10^{-5}. The largest observed scales have just entered the horizon, long after the decoupling of photons from baryons; *i.e.* causal physics have just begun to operate on them. Before decoupling, the wavelengths of these modes were so large that no causal physics could force deviations from smoothness to go away. After decoupling, the photons (to a good approximation) simply free stream. But the near-homogeneity of the CMB tells us that the universe was extremely homogeneous at the time of decoupling on a scale encompassing many regions that a priori are causally independent. How is this possible? This is a profound problem that we've glossed over by simply *assuming* that the temperature is uniform and perturbations about the zero order temperature are small.

To be more specific, remember the definition (1.6.1) of the conformal time η, which is also the causal (or comoving) horizon, characterizing the distance travelled by light since $t = 0$. We can rewrite this in the following useful form:

$$\boxed{\eta \equiv \int_0^a \frac{da'}{Ha'^2} = \int_0^a d[\ln a'] \left(\frac{1}{a'H}\right)} \, . \tag{1.8.7}$$

During the standard cosmological expansion the increasing comoving Hubble radius, $\frac{1}{aH}$, is therefore associated with an increasing comoving horizon[3], η. As we saw previously, for RD and MD universes we find

$$\begin{aligned} \eta &= \frac{1}{H_0}\int_0^a \frac{da'}{\Omega^{1/2}(a')a'^2} \\ &\propto \begin{cases} a & \text{RD} \\ a^{1/2} & \text{MD} \end{cases} . \end{aligned} \tag{1.8.8}$$

This means that the comoving horizon grows monotonically with time (at least in an expanding universe) which implies that comoving scales entering the horizon today have been far outside the horizon at CMB decoupling. Why is the CMB uniform on large scales (of order the present horizon)?

EXERCISE: Calculate the angular size of the comoving horizon at the last scattering surface at $z \simeq 1100$ as projected on to our present CMB sky, assuming a flat FRW cosmology, and that the universe contains only radiation and matter. You will need to calculate the angular diameter distance *to* last scattering and the particle horizon *at* last scattering (since proper distances, not comoving distances, appear in the derivation of the angular diameter distance).

---

[3] This justifies the common practice of often using the terms 'comoving Hubble radius' and 'comoving horizon' interchangeably. Although these terms should conceptually be clearly distinguished, this inaccurate use of language has become standard.

4. On the Problem of Initial Conditions

The flatness and horizon problems are *not* strict inconsistencies in the standard cosmological model. If one assumes that the initial value of Ω was extremely close to unity and that the universe began homogeneously (but with just the right level of inhomogeneity to explain structure formation) then the universe will continue to evolve homogeneously in agreement with observations. The flatness and horizon problems are therefore really just severe shortcomings in the predictive power of the Big Bang model. The dramatic flatness of the early universe cannot be predicted by the standard model, but must instead be assumed in the initial conditions. Likewise, the striking large-scale homogeneity of the universe is not explained or predicted by the model, but instead must simply be assumed.

5. What got the Big Bang going?

Recall the second Friedmann equation,

$$\ddot{a} = -\frac{4\pi G a}{3}(\rho + 3P) \ . \tag{1.8.9}$$

Since ordinary matter ($\rho + 3P > 0$) can only cause deceleration ($\ddot{a} < 0$), what got the Big Bang going?

B. Possible resolution of the Big Bang puzzles

1. A Crucial Idea

All Big Bang puzzles are solved by a beautifully simple idea: *invert the behavior of the comoving Hubble radius* i.e. make it *decrease* sufficiently in the very early universe (Fig. 10). The corresponding condition is that

$$\boxed{\frac{d}{dt}\left(\frac{H^{-1}}{a}\right) < 0 \quad \Rightarrow \quad \frac{d^2 a}{dt^2} > 0 \quad \Rightarrow \quad \rho + 3P < 0} \ . \tag{1.8.10}$$

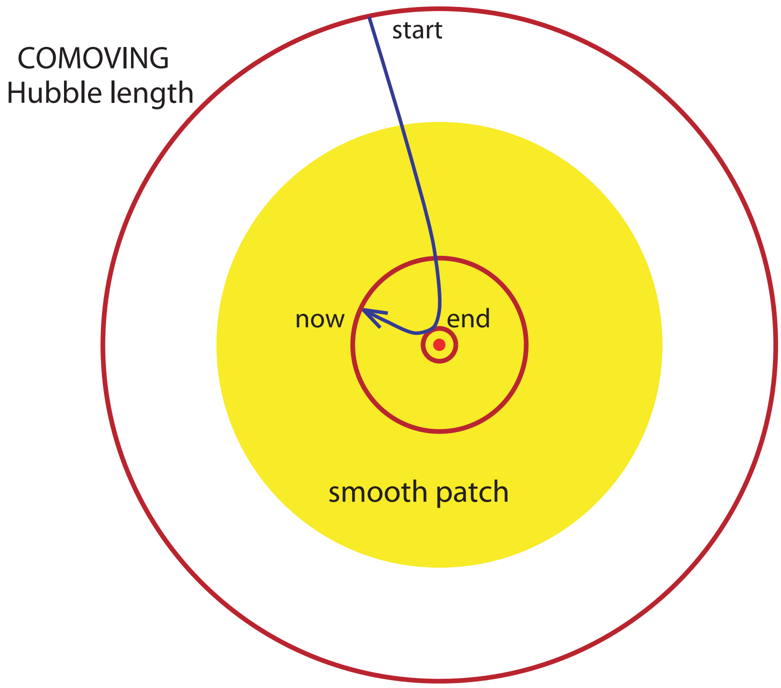

FIG. 10 Evolution of the comoving Hubble radius, $(aH)^{-1}$, in the inflationary universe. Figure credit: Adapted from Liddle and Lyth by Daniel Baumann. The comoving Hubble sphere shrinks during inflation and expands after inflation.

In a general sense, this is the essential idea about a paradigm for describing the early universe known as "inflation". The three equivalent conditions necessary for inflation are:

- Decreasing comoving horizon

$$\frac{\mathrm{d}}{dt}\left(\frac{1}{aH}\right) < 0\,. \tag{1.8.11}$$

- Accelerated expansion

$$\frac{d^2a}{dt^2} > 0\,. \tag{1.8.12}$$

- Violation of the strong energy condition – negative pressure

$$P < -\frac{1}{3}\rho\,. \tag{1.8.13}$$

Inflation presents a logical way out of the previous argument by realizing that an early epoch of rapid expansion solves the horizon problem. Einstein's equation tells us what type of energy is needed in order to produce this rapid expansion, showing that *negative pressure* is required. We will consider a scalar field theory and show that negative pressure is easy to accommodate in such a theory.

2. Comoving Horizon during Inflation

Recall the definition of the comoving horizon (= conformal time) as a logarithmic integral of the comoving Hubble radius

$$\eta = \int_0^a d[\ln a']\ \frac{1}{a'H(a')}\ . \tag{1.8.14}$$

If particles are separated by comoving distances greater than aH, they cannot currently communicate. Note the subtle distinction with η: if particles are separated by comoving distances greater than η, they could *never* have communicated. If they are separated by comoving distances greater than $(aH)^{-1}$, they can't talk to eachother *now*.

Therefore it is possible that $\eta \gg (aH)^{-1}$ now, so particles can't communicate now, but were in causal contact early on. For example, this would happen if $(aH)^{-1}$ early on was much larger than it is now, so the η integral gets most of its contribution early on. Unfortunately this does not happen during the RD and MD epochs. In those epochs, $(aH)^{-1}$ increases with time `(EXERCISE: CHECK!)`, so the largest contributions come from the most recent times.

But so far, we have been *assuming* that the universe was RD all the way back to $a = 0$! This suggests a solution: perhaps early on, the universe was not dominated by matter or radiation. For at least a brief time, $(aH)^{-1}$ *decreased* dramatically. Then η would get most of its contribution from early times before the rapid expansion of the grid.

How must $a(t)$ evolve to solve the horizon problem? To give away some of the punch line, inflationary models typically operate at energy scales 10^{15} GeV or larger. We can get a qualitative answer by assuming that the universe was RD since the end of inflation, and ignoring the relatively recent MD epoch. Remembering that $H \propto a^{-2}$ during RD, the scale factor at the end of inflation is

$$\frac{a_0 H_0}{a_e H_e} = \frac{a_0}{a_e}\left(\frac{a_e}{a_0}\right)^2 = a_e\ . \tag{1.8.15}$$

If $T_e \sim 10^{15}$ GeV,

$$a_e \sim \frac{T_0}{T_e} \sim \frac{T_0}{10^{15}\ \text{GeV}} \sim 10^{-28}\ , \tag{1.8.16}$$

i.e., $(aH)^{-1}$ at the end of inflation was 10^{28} times smaller than it is today.

For inflation to solve the horizon problem, $(aH)^{-1}$ at the start of inflation was smaller than the current comoving Hubble radius, *i.e.* the largest scales observable today. Thus, during inflation, $(aH)^{-1}$ had to *decrease* by $\sim 10^{28}$. The most common way to arrange this it to set up $H \sim$ const during inflation:

$$\frac{da}{a} = Hdt, \quad a(t) = a_e e^{H(t-t_e)}, \quad t < t_e\ , \tag{1.8.17}$$

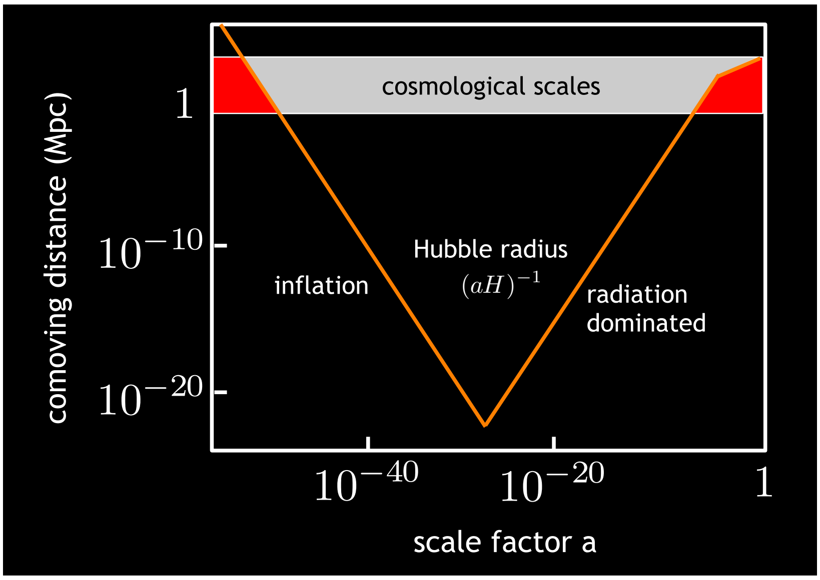

FIG. 11 Solution of the Horizon Problem. Scales of cosmological interest were larger than the Hubble radius until $a \sim 10^{-5}$. However, very early on, before inflation operated, all scales of interest were smaller than the Hubble radius and therefore susceptible to microphysical processing. Similarly, at very late time, scales of cosmological interest came back within the Hubble radius. Adapted from Dodelson.

where t_e is the time at the end of inflation. Thus, decreasing $(aH)^{-1}$ during inflation is solely due to the exponential increase in $a(t)$. For the scale factor to increase by 10^{28}, $H(t - t_e)$ must be $\sim \ln(10^{28}) \sim 64$.

`EXERCISE:` Assuming a_e corresponded to 10^{15} GeV, compute the number of "e–folds" of inflation needed to solve the horizon problem, accounting for the RD/MD transition at $a_{\rm EQ}$.

Notice the symmetry of the inflationary solution (*cf.* Fig. 11). Scales just entering the horizon today, ~ 60 e-folds *after* the end of inflation, left the horizon ~ 60 e-folds *before* the end of inflation.

So far, we have discussed inflation in terms of comoving coordinates. But it is also profitable to think of exponential expansion in terms of physical coordinates. The physical size of a causally connected region exponentially increases during inflation. So regions that we observe to be cosmological today were actually microscopically small before inflation, and in causal contact.

The total comoving horizon is not an effective time parameter after inflation, because it becomes very large early on, then changes very little during RD and MD. We can rectify this by subtracting off the primordial part

$$\eta = \int_{t_e}^{t} \frac{dt'}{a(t')} \ , \tag{1.8.18}$$

so that the total comoving horizon is $\eta_{\rm prim} + \eta$. So during inflation, η is negative, but monotonically increasing. A scale leaves the horizon when $k|\eta| < 1$, and reenters the horizon when $k|\eta| > 1$.

3. Flatness Problem Revisited

Recall the Friedmann equation in the form

$$|1 - \Omega(a)| = \frac{\kappa}{(aH)^2} \, . \tag{1.8.19}$$

Inflation ($H \approx$ const., $a = e^{Ht}$) is characterized by a decreasing comoving horizon which *drives the universe toward flatness* (rather than away from it),

$$|1 - \Omega(a)| \propto \frac{1}{a^2} = e^{-2Ht} \to 0 \quad \text{as} \quad t \to \infty\,. \tag{1.8.20}$$

This solves the flatness problem! $\Omega = 1$ is an attractor during inflation.

`EXERCISE:` Assume $\Omega_0 \sim 0.3$ today, and $\Omega_\Lambda = 0$. Extrapolate $\Omega(t) - 1$ back to the end of inflation, then through 60 e-folds of inflation. What is $\Omega(t) - 1$ right before the 60 e-folds of inflation?

---

4. Horizon Problem Revisited

A decreasing comoving horizon means that large scales entering the present horizon were inside the horizon before inflation (see Fig. 11). Causal physics before inflation therefore established thermal equilibrium and spatial homogeneity. The uniformity of the CMB is not a mystery.

5. Origin of primordial inhomogeneities

Besides solving the Big Bang puzzles the decreasing comoving horizon during inflation is the key feature required for the quantum generation of cosmological perturbations described in the third part of the course. During inflation, quantum fluctuations are generated on subhorizon scales, but exit the horizon once the Hubble radius becomes smaller than their comoving wavelength. In physical coordinates this corresponds to the superluminal expansion stretching perturbations to acausal distances. They become classical superhorizon density perturbations which reenter the horizon in the subsequent Big Bang evolution and then undergo gravitational collapse to form the large scale structure in the universe.

6. Conformal Diagram of Inflationary Cosmology

A truly illuminating way of visualizing inflation is with the aid of a conformal spacetime diagram. Reminder: the flat FRW metric in conformal time η becomes

$$ds^2 = a^2(\eta)\left[d\eta^2 - d\vec{x}^2\right]\,, \tag{1.8.21}$$

and during matter or radiation domination the scale factor evolves as

$$a(\eta) \propto \begin{cases} \eta & \text{RD} \\ \eta^2 & \text{MD}\,. \end{cases} \tag{1.8.22}$$

If the universe had always been dominated by matter or radiation, this would imply the existence of the *big bang singularity* at $\eta = 0$,

$$a(\eta \equiv 0) = 0\,. \tag{1.8.23}$$

The conformal diagram corresponding to standard Big Bang cosmology is given in Fig. 12. The horizon problem is apparent. Each spacetime point in the conformal diagram has an associated past light cone which defines its causal past. Two points on a given $\eta =$ constant surface are in causal contact if their past light cones intersect at the Big Bang, $\eta = 0$. This means that the surface of last scattering ($\eta_{\rm CMB}$) consisted of many causally disconnected regions that won't be in thermal equilibrium. The uniformity of the CMB on large scales hence becomes a serious puzzle.

In de Sitter space, the scale factor is

$$a(\eta) = -\frac{1}{H\eta}\,, \tag{1.8.24}$$

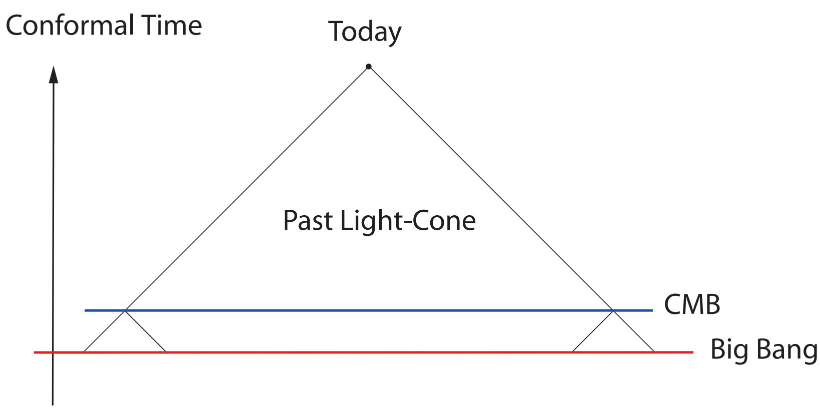

FIG. 12 Conformal diagram of Big Bang cosmology. The CMB at last scattering consists of 10^5 causally disconnected regions. Figure credit: Daniel Baumann.

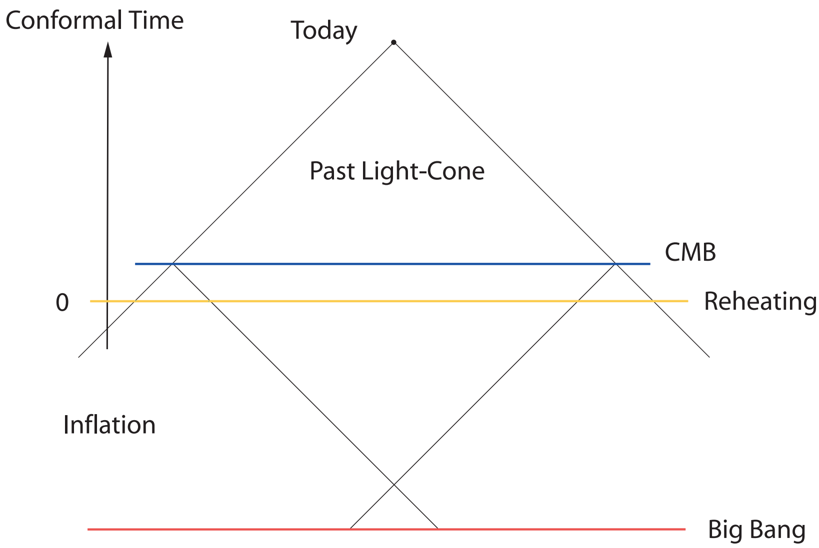

FIG. 13 Conformal diagram of inflationary cosmology. Inflation extends conformal time to negative values. The end of inflation creates an "apparent" Big Bang at $\eta = 0$. There is, however, no singularity at $\eta = 0$ and the light cones intersect at an earlier time iff inflation lasts for at least 60 e-folds. Figure credit: Daniel Baumann.

and the singularity, $a = 0$, is pushed to the infinite past, $\eta \to -\infty$. The scale factor (1.8.24) becomes infinite at $\eta = 0$! This is because we have assumed de Sitter space with $H =$ const., which means that inflation will continue forever with $\eta = 0$ corresponding to the infinite future $t \to +\infty$. In reality, inflation ends at some finite time, and the approximation (1.8.24) although valid at early times, breaks down near the end of inflation. So the surface $\eta = 0$ is not the Big Bang, but the end of inflation. The initial singularity has been pushed back arbitrarily far in conformal time $\eta \ll 0$, and light cones can extend through the apparent Big Bang so that apparently disconnected points are in causal contact. This is summarized in the conformal diagram in Fig. 13.

IX. INFLATION: IMPLEMENTATION WITH A SCALAR FIELD

Inflation, an epoch in which the universe accelerates, solves a number of puzzles associated with the standard Big Bang cosmology. During a phase of accelerated expansion, H^{-1} (the physical Hubble radius) remains fixed. So particles initially in causal contact can no longer communicate. Regions separated by cosmological distances today were actually in causal contact before/during inflation. At that time, these regions were given the necessary initial conditions, smoothness, and *small perturbations about smoothness* we observe today. The last part of the course will detail how these small perturbations came about.

In this lecture, we will consider how to implement inflation using a scalar field with special dynamics. Although no fundamental scalar field has yet been detected in experiments, there are fortunately plenty of such fields in theories beyond the standard model of particle physics. In fact, in string theory, for example, there are numerous scalar fields (moduli), although it proves very challenging to find just one with the right characteristics to serve as an inflaton candidate. In the following we will therefore describe the dynamics of a generic scalar field leaving the connection with fundamental particle theory for a future revolution in theoretical physics. We will leave the physics of the generation of primordial perturbations during inflation to the discussion of perturbation theory later in the course, only dealing with them in a heuristic way to describe observables of this theory.

A. Negative pressure

From the second Friedmann equation (1.4.26), an accelerating universe requires

$$\ddot{a} > 0 \Rightarrow (\rho + 3P) < 0 \Rightarrow P < -\frac{\rho}{3} \Rightarrow w < -\frac{1}{3} \, . \tag{1.9.1}$$

Since $\rho \geq 0$, this implies *negative pressure*, $P < 0$. This can't be ordinary matter or radiation!

B. Implementation with a scalar field

Can a scalar field have $\rho + 3P < 0$? We can rewrite the energy momentum tensor for a homogeneous scalar field ϕ, (1.7.39) in the mixed form,

$$T^{\mu}_{\ \nu} = g^{\mu\beta}\phi_{,\beta}\phi_{,\nu} - \delta^{\mu}_{\ \nu}\left[\frac{1}{2}g^{\lambda\alpha}\phi_{,\alpha}\phi_{,\lambda} - V(\phi)\right] \, , \tag{1.9.2}$$

and equate it to (1.4.9). Here, $V(\phi)$ is the potential, *e.g.*, for a free field with mass m, $V(\phi) = \frac{1}{2}m^2\phi^2$. For a homogeneous field, only time-derivatives are relevant. Let us assume $\phi(x,t) = \phi^{(0)}(t) + \delta\phi(\vec{x},t)$ and derive the density, pressure, and time-evolution of the homogeneous part. Equating the time-time component, we find:

$$\boxed{\rho_\phi = \underbrace{\frac{1}{2}\left(\frac{\partial\phi^{(0)}}{\partial t}\right)^2}_{\text{kinetic energy}} + \underbrace{V(\phi^{(0)})}_{\text{potential energy}}} \, , \tag{1.9.3}$$

which looks like the energy density of a single particle with position $\phi^{(0)}$ moving in a potential $V(\phi)$. Equating the space-space component, we find that:

$$\boxed{P_\phi = \frac{1}{2}\left(\frac{\partial\phi^{(0)}}{\partial t}\right)^2 - V(\phi^{(0)})} \, . \tag{1.9.4}$$

Thus, for negative pressure,

$$P < 0 \Longrightarrow \underbrace{V(\phi^{(0)}) > \frac{1}{2}\left(\frac{\partial\phi^{(0)}}{\partial t}\right)^2}_{\text{PE>KE}} \, . \tag{1.9.5}$$

The resulting equation of state

$$w = \frac{\frac{1}{2}\dot{\phi}^2 - V}{\frac{1}{2}\dot{\phi}^2 + V} \tag{1.9.6}$$

shows that a scalar field can lead to negative pressure ($w < 0$) and accelerated expansion ($w < -1/3$) if the potential energy V dominates over the kinetic energy $\frac{1}{2}\dot{\phi}^2$.

As an example, consider a field trapped in a false vacuum. The energy density is constant since $\phi^{(0)}$ is a constant. Thus,

$$\frac{\dot{a}}{a} = \sqrt{\frac{8\pi G\rho}{3}} = \text{const} . \tag{1.9.7}$$

Thus, we obtain exponential expansion with $H \propto \rho^{1/2} = \text{const}$:

$$\eta_{\text{prim}} = \frac{1}{H_e a_e}\left[e^{H(t_e - t_b)} - 1\right] , \tag{1.9.8}$$

where t_b and t_e denote the beginning and end of inflation, respectively. So if the field is trapped for 60 e-foldings, $H(t_e - t_b) > 60$, the horizon problem is solved. This is actually "Old Inflation", the original Guth (1981) formulation. It doesn't work because "bubbles" of localized regions tunnelling into the true vacuum never coalesce in order for the universe as a whole to move to the true vacuum. In "New Inflation", Linde, Steinhardt and Albrecht postulated that the scalar field rolls slowly toward the true vacuum in a nearly flat potential, where $\rho \sim \text{const}$.

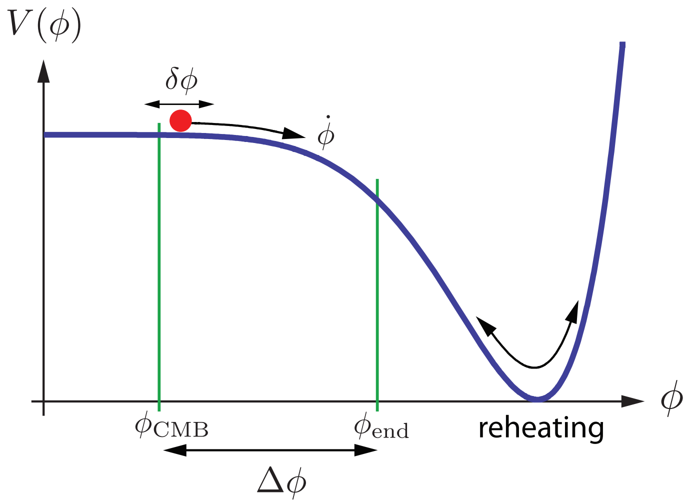

FIG. 14 Example of an inflaton potential. Acceleration occurs when the potential energy of the field, $V(\phi)$, dominates over its kinetic energy, $\frac{1}{2}\dot{\phi}^2$. Inflation ends at ϕ_{end} when the kinetic energy has grown to become comparable to the potential energy, $\frac{1}{2}\dot{\phi}^2 \approx V$. CMB fluctuations are created by quantum fluctuations $\delta\phi$ about 60 e-folds before the end of inflation. At reheating, the energy density of the inflaton is converted into radiation. Figure credit: Daniel Baumann.

What is the equation of motion of a scalar field in an FRW background? Reminding ourselves of the scalar field action of Eq. (1.7.27), for a scalar field minimally coupled to gravity, the action is

$$S_\phi^{(\text{minimally}-\text{coupled})} = \int d^4x\sqrt{-g}\left[-\frac{R}{16\pi G} + \frac{1}{2}g^{\mu\nu}(\nabla_\mu\phi)(\nabla_\nu\phi) - V(\phi)\right] . \tag{1.9.9}$$

Hence, reminding ourselves of Eq. (1.7.28), the equation of motion for the homogeneous field ($\partial_i\phi = 0$) is

$$g^{\mu\nu}\nabla_\mu(\partial_\nu\phi) + \frac{dV}{d\phi} = 0 . \tag{1.9.10}$$

Noting the presence of the covariant derivative and using the Christoffel symbols (1.3.13) for the flat FRW cosmology, the equation of motion can be written as:

$$\boxed{\ddot{\phi}^{(0)} + 3H\dot{\phi}^{(0)} + \frac{dV(\phi^{(0)})}{d\phi} = 0} , \tag{1.9.11}$$

with Hubble parameter

$$\boxed{H^2 = \frac{8\pi G}{3}\left[\frac{1}{2}\left(\dot{\phi}^{(0)}\right)^2 + V(\phi^{(0)})\right]}\,. \tag{1.9.12}$$

`EXERCISE:` Derive Eq. (1.9.11).

By directly varying the action with respect to $\delta\phi$, (*cf.* Eq. (1.7.17)) the equation of motion can also be written in the following useful form:

$$\boxed{\frac{\delta S}{\delta\phi} = \frac{1}{\sqrt{-g}}\partial_\mu(\sqrt{-g}\partial^\mu\phi) + \frac{dV(\phi)}{d\phi} = 0}\,, \tag{1.9.13}$$

where in the flat FRW background, $\sqrt{-g} = a^3(t)$.

`EXERCISE:` Obtain Eq. (1.9.11) from yet another route by appropriately combining the Friedmann equations for a scalar-field dominated universe. By changing the independent variable to the conformal time η, show that the equation of motion can be written in the following useful form:

$$\boxed{\frac{d^2\phi^{(0)}}{d\eta^2} + 2aH\frac{d\phi^{(0)}}{d\eta} + a^2\frac{dV(\phi^{(0)})}{d\phi} = 0}\,. \tag{1.9.14}$$

C. Slow-Roll Inflation

$\boxed{\text{In the following discussion, for notational convenience we will set } 8\pi G \equiv M_{\rm Pl}^{-2} \equiv 1.}$

Inflation occurs if the field is evolving slow enough that the potential energy dominates over the kinetic energy, and the second time derivative of ϕ is small enough to allow this slow-roll condition to be maintained for a sufficient period. Thus, inflation requires

$$\dot{\phi}^2 \ll V(\phi) \tag{1.9.15}$$

$$|\ddot{\phi}| \ll |3H\dot{\phi}|, |V_{,\phi}|\,. \tag{1.9.16}$$

Satisfying these conditions requires the smallness of two dimensionless quantities known as *slow-roll parameters*

$$\epsilon_V(\phi) \equiv \frac{1}{2}\left(\frac{V_{,\phi}}{V}\right)^2 \tag{1.9.17}$$

$$\eta_V(\phi) \equiv \frac{V_{,\phi\phi}}{V}\,. \tag{1.9.18}$$

In the slow-roll regime

$$\epsilon_V, |\eta_V| \ll 1 \tag{1.9.19}$$

the background evolution is

$$H^2 \approx \frac{1}{3}V(\phi) \approx \text{const}\,, \tag{1.9.20}$$

$$\dot{\phi} \approx -\frac{V_{,\phi}}{3H}\,, \tag{1.9.21}$$

and the spacetime is approximately *de Sitter*

$$a(t) \sim e^{Ht}\,, \quad H \approx \text{const}\,. \tag{1.9.22}$$

EXERCISE: To understand the relation between the slow-roll condition (1.9.19) and inflation convince yourself that the following is true

$$\frac{\ddot{a}}{a} = H^2 \left(1 + \frac{\dot{H}}{H^2}\right) \approx H^2(1-\epsilon_V)\,. \tag{1.9.23}$$

Consequently, if the slow-roll approximation is valid ($\epsilon_V \ll 1$), then inflation is guaranteed. However, this condition is sufficient but not necessary since the validity of the slow-roll approximation was required to establish the second equality in (1.9.23).

Inflation ends when the slow-roll conditions (1.9.19) are violated[4]

$$\epsilon_V(\phi_{\rm end}) \approx 1\,. \tag{1.9.24}$$

The number of e-folds before inflation ends is

$$\begin{aligned} N(\phi) &\equiv \ln \frac{a_{\rm end}}{a} \\ &= \int_t^{t_{\rm end}} H dt \\ &\approx \int_{\phi_{\rm end}}^{\phi} \frac{V}{V_{,\phi}} d\phi\,. \end{aligned} \tag{1.9.25}$$

D. What is the Physics of Inflation?

The inflationary proposal requires a huge extrapolation of the known laws of physics. It is therefore not surprising that the physics governing this phase of superluminal expansion is still very uncertain. In the absence of a complete theory standard practice has been a phenomenological approach, where an effective potential $V(\phi)$ is postulated. Ultimately, $V(\phi)$ has to be derived from a fundamental theory.

Understanding the (micro)physics of inflation remains one of the most important open problems in modern cosmology and theoretical physics. Explicit particle physics models of inflation remain elusive. A natural microscopic explanation for inflation has yet to be uncovered. Nevertheless, there have recently been interesting efforts to derive inflation from string theory. Inflation in string theory is still in its infancy, but it seems clear that our understanding of inflation will benefit greatly from a better understanding of moduli stabilization and supersymmetry breaking in string theory. Hopefully, this will give some insights into which models of inflation are microscopically viable, meaning that they can be derived from explicit string compactifications. Given the prospect of explicit and controllable models of inflation in string theory, one is led to ask whether these theories have specific observational signatures. In particular, it will be interesting to explore whether there are predictions that while non-generic in effective field theory, may still have a well-motivated microscopic origin in string theory.

---

[4] This can be made exact by the use of Hubble slow-roll parameters instead of the potential slow-roll parameters we introduced here (see § IX.E).

E. Slow-Roll Inflation in the Hamilton-Jacobi Approach (NON–EXAMINABLE)

In the following discussion, for notational convenience we will set $8\pi G \equiv M_{\rm Pl}^{-2} \equiv 1$.

1. Hamilton-Jacobi Formalism

The Hamilton-Jacobi approach treats the Hubble expansion rate $H(\phi) = \mathcal{H}/a$ as the fundamental quantity, considered as a function of time. Consider

$$H_{,\phi} = \frac{H'}{\phi'} = \frac{-\frac{1}{a}(\mathcal{H}^2 - \mathcal{H}')}{\phi'} = -\frac{\phi'}{2a}\,, \tag{1.9.26}$$

where we used $\mathcal{H}^2 - \mathcal{H}' = a^2(\rho + P)/2 = (\phi')^2/2$ and primes are derivatives with respect to conformal time. This gives the master equation

$$\boxed{\frac{d\phi}{dt} = \frac{\phi'}{a} = -2H_{,\phi}}\,. \tag{1.9.27}$$

This allows us to rewrite the Friedmann equation

$$H^2 = \frac{1}{3}\left[\frac{1}{2}\left(\frac{d\phi}{dt}\right)^2 + V(\phi)\right] \tag{1.9.28}$$

in the following way

$$\boxed{[H_{,\phi}]^2 - \frac{3}{2}H^2 = \frac{1}{2}V(\phi)}\,. \tag{1.9.29}$$

Notice the following important consequence of the Hamilton-Jacobi equation (1.9.29): For any specified function $H(\phi)$, it produces a potential $V(\phi)$ which admits the given $H(\phi)$ as an exact inflationary solution. Integrating (1.9.27)

$$\int dt = -\frac{1}{2}\int \frac{d\phi}{H'(\phi)} \tag{1.9.30}$$

relates ϕ to proper time t. This enables us to obtain $H(t)$, which can be integrated to give $a(t)$. The Hamilton-Jacobi formalism can be used to generate infinitely many inflationary models with exactly known analytic solutions for the background expansion. Here we are more concerned with the fact that it allows an elegant and intuitive definition of the slow-roll parameters.

2. Hubble Slow-Roll Parameters

During slow-roll inflation the background spacetime is approximately de Sitter. Any deviation of the background equation of state

$$w = \frac{P}{\rho} = \frac{(\phi')^2/2a^2 - V}{(\phi')^2/2a^2 + V}$$

from the perfect de Sitter limit $w = -1$ may be defined by the parameter

$$\boxed{\epsilon_H \equiv \frac{3}{2}(1+w)}\,. \tag{1.9.31}$$

We can express the Friedmann equations

$$\mathcal{H}^2 = \frac{1}{3}a^2\rho \tag{1.9.32}$$

$$\mathcal{H}' = -\frac{1}{6}a^2(\rho + 3P) \tag{1.9.33}$$

in terms of ϵ_H

$$\mathcal{H}^2 = \frac{1}{3}\frac{(\phi')^2}{\epsilon_H} \tag{1.9.34}$$

$$\mathcal{H}' = \mathcal{H}^2(1-\epsilon_H)\,. \tag{1.9.35}$$

Hence,

$$\epsilon_H = 1 - \frac{\mathcal{H}'}{\mathcal{H}^2} = \frac{d(H^{-1})}{dt}\,. \tag{1.9.36}$$

Note that this can be interpreted as the rate ot change of the Hubble parameter during inflation H with respect to the number of e-foldings $dN = H dt = -\frac{1}{2}\frac{H(\phi)}{H_{,\phi}}\mathrm{d}\phi$

$$\boxed{\epsilon_H = -\frac{d[\ln H]}{dN} = 2\left(\frac{H_{,\phi}}{H}\right)^2}\,. \tag{1.9.37}$$

Analogously we define the second slow-roll parameter as the rate of change of $H_{,\phi}$

$$\boxed{\eta_H = -\frac{d[\ln|H_{,\phi}|]}{dN} = 2\frac{H_{,\phi\phi}}{H}} \tag{1.9.38}$$

Using the Hamilton-Jacobi master equation (1.9.27) this is also

$$\eta_H = \frac{d[\ln|\mathrm{d}\phi/dt|]}{dN}\,. \tag{1.9.39}$$

3. Slow-Roll Inflation

By definition, slow-roll corresponds to a regime where all dynamical characteristics of the universe, measured in physical (proper) units, change little over a single e-folding of expansion. This ensures that the primordial perturbations are generated with approximately equal power on all scales, leading to a scale-invariant perturbation spectrum.

Since ϵ_H and η_H characterize the rate of change of H and $H_{,\phi}$ with e-foldings, slow-roll is naturally defined by

$$\epsilon_H \ll 1 \tag{1.9.40}$$

$$|\eta_H| \ll 1\,. \tag{1.9.41}$$

The first slow-roll condition implies

$$\epsilon_H \ll 1 \quad \Rightarrow \quad \mathcal{H}^2 = \frac{1}{3}\frac{(\phi')^2}{\epsilon_H} \gg (\phi')^2\,, \tag{1.9.42}$$

so that the slow-roll limit of the first Friedmann equation is

$$\mathcal{H}^2 \approx \frac{1}{3}a^2 V\,. \tag{1.9.43}$$

The second slow-roll condition implies

$$\eta_H = \frac{d[\ln|\mathrm{d}\phi/dt|]}{dN} = \frac{1}{H|\mathrm{d}\phi/dt|}\frac{\mathrm{d}^2\phi}{dt^2} \ll 1 \qquad \Rightarrow \qquad |\mathrm{d}^2\phi/dt^2| \ll H|d\phi/dt| \tag{1.9.44}$$

so that the Klein-Gordon equation reduces to

$$\dot{\phi} \approx -\frac{a^2 V'}{3\mathcal{H}}\,. \tag{1.9.45}$$

In section IX.C we defined a second set of common slow-roll parameters in terms of the local shape of the potential $V(\phi)$

$$\epsilon_V \equiv \frac{1}{2}\left(\frac{V_{,\phi}}{V}\right)^2 \tag{1.9.46}$$

$$\eta_V \equiv \frac{V_{,\phi\phi}}{V}\,. \tag{1.9.47}$$

$\epsilon_H(\phi_{\rm end}) \equiv 1$ is an exact definition of the end of inflation, while $\epsilon_V(\phi_{\rm end}) = 1$ is only an approximation. In the slow-roll regime the following relations hold

$$\epsilon_H \approx \epsilon_V \tag{1.9.48}$$

$$\eta_H \approx \eta_V - \epsilon_V \,. \tag{1.9.49}$$