# Lecture II: Basics of the homogeneous, isotropic Universe

(Dated: January 11, 2019)

## I. INTRODUCTION

This lecture will introduce the physics of expanding Universes. We will cover:

- *Motivation* – Why do we think the Universe is homogeneous and isotropic?
- *Spatial geometry* – What are the possible geometrical structures of the Universe?
- *Expansion kinematics* – How do we describe the expansion of the Universe?
- *Expansion dynamics* – The Friedmann equations, the basic dynamical equations controlling the expansion rate of the Universe (Lecture II).

For the discussion of cosmic expansion, we will need to use relativity, and so we will use units where $c = 1$. I will put the factor of $c$ back later if it is convenient (e.g., we need to do a calculation in SI units).

## II. THE HOMOGENEOUS, ISOTROPIC UNIVERSE

We live in a Universe that, on everyday scales, is very inhomogeneous: the matter we see around us is clumped into planets, stars, galaxies, clusters of galaxies, etc. The idea that on very large scales, the Universe is homogeneous (the same at all *positions*) and isotropic (the same in all *directions*) is often called the **Cosmological Principle**.

While initially a philosophically motivated idea, the Cosmological Principle has held up well when confronted with empirical evidence. Isotropy is the easiest to probe:

- The intensity of the cosmic microwave background radiation varies by only a few parts in $10^5$ in different directions.
- The large-scale distribution of distant quasars on the sky is isotropic with variations of at most of order 1% (limited by observational uncertainties).

Of course, there are real variations in the microwave background, and quasars exist at specific locations. The sense in which the Cosmological Principle is valid, then, is twofold: one is a statement about the structure of the Universe on large scales; and second is a statement that individual objects (e.g. quasars) are distributed according to a statistical distribution that does not favor any point or direction. We will explore both senses of the cosmological principle throughout the course.

Homogeneity is inherently more difficult to test than isotropy – we can point a telescope in a different direction, but we can't (yet) go to another galaxy to observe the Universe. If the Universe were isotropic as seen from *every* point, then it must be homogeneous, but all we can directly observe is the isotropy as seen from the Solar System. Therefore all of our tests of homogeneity are *indirect*:

- You might imagine using counts of galaxies – are there 8 times as many galaxies of a given type if you search twice as far? But such tests, in their simplest form, are limited to the local Universe, since if you look farther away you are seeing an earlier epoch of cosmic evolution, and hence are looking at inherently different objects.
- The most powerful such tests come from the cosmic microwave background. If we lived in a large-scale over-density, then all of the galaxies would be falling toward us, and they would up-scatter CMB photons to higher frequencies. This would distort the spectrum of the CMB, making it no longer a perfect blackbody.

Nevertheless, our constraints on cosmic inhomogeneity are weaker than those on anisotropy.

## III. POSSIBLE SPATIAL GEOMETRIES

We now turn our attention to the possible spatial geometrical structure of the Universe. Our basic question is:

*What are the 3-dimensional surfaces that have no preferred point (are homogeneous) and no preferred direction (are isotropic)?*

## A. Flat geometry

You learned about one of the possibilities in school – the Euclidean or **flat** geometry, $\mathbb{R}^3$. We can describe it by writing the geometry in terms of a **line element**: the law that describes the distance between any two infinitesimally separated points. (You will also see this described as a **metric**; technically a metric is a formula for the dot product of two vectors, but it contains the same information as the line element and so to a physicist they are equivalent.) The Euclidean geometry line element is simply the Pythagorean theorem:

$$ds^2 = dx^2 + dy^2 + dz^2. \tag{1}$$

In observational cosmology, where we observe from one particular point, it is convenient to write this in spherical polar coordinates instead. The line element is then:

$$ds^2 = dr^2 + (r\,d\theta)^2 + (r\sin\theta\,d\phi)^2 = dr^2 + r^2(d\theta^2 + \sin^2\theta\,d\phi^2). \tag{2}$$

Here the distance from the origin is $r$, the polar angle or co-latitude is $\theta$, and the longitude is $\phi$. We've written this equation as well using the Pythagorean theorem, the only difference being that in spherical polar coordinates we need to remember that (i) the $r$, $\theta$, and $\phi$ directions are all orthogonal; and (ii) the arc length in the angular directions ($r\,d\theta$ or $r\sin\theta\,d\phi$) is not the same as the angle traversed, but differs by a factor of the radius of the circle ($r$ for the great circle in the latitude direction, $r\sin\theta$ for the parallel of constant latitude). The coordinates span the range:

$$r \geq 0; \quad 0 \leq \theta \leq \pi; \quad \text{and} \quad 0 \leq \phi < 2\pi. \tag{3}$$

*Notes* — You will get more practice with line elements in the homework, and we will work a few examples in class so that you can get comfortable. In general, the line element $ds^2$ is quadratic in the infinitesimal coordinate displacements, so for example in 2 dimensions we could write it as

$$ds^2 = E(x,y)\,dx^2 + 2F(x,y)\,dx\,dy + G(x,y)\,dy^2. \tag{4}$$

In $N$ dimensions, there are $N(N+1)/2$ functions of the coordinates that determine the line element. In most of the coordinate systems you are familiar with, the coordinates are **orthogonal** (e.g. in spherical polar coordinates, the $r$, $\theta$, and $\phi$ directions are orthogonal) and there aren't any mixed terms (e.g. $dr\,d\theta$). We won't need to use non-orthogonal coordinate systems in this class, except maybe at the very end, but you should be aware of their existence.

## B. Closed geometry

Just because Euclidean geometry is homogeneous and isotropic doesn't mean it's the only possibility. The surface of a 2-sphere of radius $R$ is an example of a 2-dimensional surface embedded in 3-dimensional space that is homogeneous (all of the points on the surface of the sphere are equivalent) and isotropic (you can sit at any point, draw an arrow tangent to the surface of the sphere in any direction, and those directions are also equivalent). It's described by the line element:

$$ds^2 = (R\,d\theta)^2 + (R\sin\theta\,d\phi)^2. \tag{5}$$

By analogy, we can write the geometry for a 3-sphere $S^3$ of radius $R$, which could be embedded in 4-dimensional space. We need a new polar coordinate $\alpha$ (spanning the range from 0 to $\pi$):

$$ds^2 = (R\,d\alpha)^2 + (R\sin\alpha\,d\theta)^2 + (R\sin\alpha\sin\theta\,d\phi)^2 \tag{6}$$

Remember that while this is a sphere in 4-dimensional space, from the perspective of the geometry only the 3-dimensional surface is physically meaningful; the space in which it is embedded is not accessible and is only a mathematical construction that we use to show that the space is homogeneous and isotropic.

The 3-sphere is called a **closed** geometry.

The radial coordinate written here is the polar angle $\alpha$. Two other choices of coordinate are the radial coordinate $\chi = R\alpha$ and the shell radius $r = R\sin\alpha$. The closed geometry can be written in terms of either:

$$\begin{aligned} ds^2 &= R^2 d\alpha^2 + (R\sin\alpha)^2(d\theta^2 + \sin^2\theta\,d\phi^2) \\ &= d\chi^2 + \left(R\sin\frac{\chi}{R}\right)^2(d\theta^2 + \sin^2\theta\,d\phi^2) \\ &= \frac{dr^2}{1-Kr^2} + r^2(d\theta^2 + \sin^2\theta\,d\phi^2), \end{aligned} \tag{7}$$

where in the last line we have defined the **spatial curvature** $K = 1/R^2$. Here $\chi$ is the distance from the North Pole to the point of interest, which makes the radial part of the line element look simple; and $r$ is the radius (in the embedding space) of the spherical shell of constant hyper-co-latitude $\alpha$.

The coordinate $\chi$ ranges over $0 \le \chi \le \pi R$. The coordinate $r$ ranges from $0 \le r \le R$, but is ambiguous in the sense that it only covers the Northern Hemisphere; the Southern Hemisphere repeats the same values of $r$.

An interesting aspect of the closed geometry is that it has finite volume. The volume element is given by multiplying the line element in the three orthogonal directions:

$$d^3V = (R\,d\alpha)(R\sin\alpha\,d\theta)(R\sin\alpha\sin\theta\,d\phi) = R^3\sin^2\alpha\sin\theta\,d\alpha\,d\theta\,d\phi \tag{8}$$

and the total volume is

$$V_{\rm tot} = R^3\int_0^\pi \sin^2\alpha\,d\alpha\int_0^\pi \sin\theta\,d\theta\int_0^{2\pi} d\phi = R^3\left(\frac{\pi}{2}\right)(2)(2\pi) = 2\pi^2R^3. \tag{9}$$

### C. Open geometry

Once we accept that in closed geometry the 4-dimensional Euclidean space in which the 3-sphere was embedded is fictitious, we are free to consider even more bizarre possibilities. In particular, in Eq. (7), there is no longer any need to force the curvature $K$ to be positive – it might as well be negative. This leads to **open geometry**, $\mathbb{H}^3$. Since $K = 1/R^2$, the radius of curvature $R$ is imaginary, but as long as the coordinates we use and the line element in those coordinates are real, there are no problems. Writing $R = iS$, we see that the coordinates $\chi$ and $r$ are related by

$$r = R\sin\frac{\chi}{R} = iS\sin\frac{\chi}{iR} = iS\left(\frac{\chi}{iS} - \frac{\chi^3}{3!i^3S^3} + \frac{\chi^5}{5!i^5S^5} - ...\right) = S\left(\frac{\chi}{S} + \frac{\chi^3}{3!S^3} + \frac{\chi^5}{5!S^5} + ...\right) = S\sinh\frac{\chi}{S}. \tag{10}$$

The line element is

$$\begin{aligned} ds^2 &= d\chi^2 + \left(S\sinh\frac{\chi}{S}\right)^2(d\theta^2 + \sin^2\theta\,d\phi^2) \\ &= \frac{dr^2}{1-Kr^2} + r^2(d\theta^2 + \sin^2\theta\,d\phi^2), \end{aligned} \tag{11}$$

with $K < 0$. By analytic continuation, homogeneity and isotropy are preserved.

Unlike the closed case, where $r$ had a maximum value of $R = K^{-1/2}$, here $r$ and $\chi$ have no upper limits, and the volume is infinite. In fact, the volume element is

$$d^3V = (d\chi)(S\sinh\frac{\chi}{S}\,d\theta)(S\sinh\frac{\chi}{S}\sin\theta\,d\phi) = S^2\sin^2\frac{\chi}{S}\sin\theta\,d\chi\,d\theta\,d\phi \tag{12}$$

and the volume within distance $\chi$ of the origin is

$$V(<\chi) = S^2\int_0^\chi \sin^2\frac{\chi}{S}\,d\chi\int_0^\pi \sin\theta\,d\theta\int_0^{2\pi} d\phi = \pi S^3\left(\sinh\frac{2\chi}{S} - \frac{2\chi}{S}\right), \tag{13}$$

which increases *exponentially* with $\chi$. In this sense, the open geometry is even larger than the (infinite) Euclidean geometry.

It can be shown that the flat, closed, and open models are the only possible geometries. [You may do this by brute force, allowing an arbitrary function $r(\chi)$ and demanding the ability to change coordinates to put a new point at the origin; or if you studied Lie groups you may be clever, write the generators of the translations and rotations, and consider the possible commutators $[P_i, P_j]$ of the translations.]

### D. Topology

Finally, we consider topology: for any given geometry, it is possible to construct different topologies by adjusting the legal range of the coordinates and how the "boundaries" of these regions are connected to each other. As a simple example, a circle of circumference $C$ looks geometrically like the real line, but with the range of coordinate restricted to $0 \le x < C$ and with points $x = 0$ and $x = C$ identified with each other. An observer in this 1-dimensional space will see copies of themselves in both directions at a distance $C$.

Topological identifications such as this are very restricted in higher dimensions by the principle of isotropy. If we, the observers at some position $O$ (say the North Pole or origin) look up in the sky, the nearest image of ourselves that we see will be at some distance $C$ – and this must be the same distance in every direction, i.e. they occupy a spherical shell of radius $C$. You might think this is impossible, since those images themselves must be separated by at least distance $C$. The only loophole is to consider the closed universe, and set $C = \pi R$, so that the North Pole is identified with the South Pole. This topologically reduced space is called the **real-projective space**, $\mathbb{R}P^3$. It is like $S^3$ except that each point is equivalent to the opposite or antipodal point, and hence the total volume is $\frac{1}{2}$ of that of $S^3$: $V_{\rm tot} = \pi^2 R^3$.

Other topological identifications that do not respect isotropy can be considered – in this case, we would have a space that is locally isotropic (a ball of radius $\epsilon$ around any observer is spherically symmetric) but not globally isotropic (e.g., there might be special directions we could look and see copies of ourselves). Thus far, these solutions do not appear to be relevant to the real Universe, and we won't discuss them further in this class.

## IV. EXPANSION KINEMATICS

We now turn our attention to the global structure of spacetime and the description of the kinematics of the expansion.

### A. Review: line element in special relativity

Let's first revisit how kinematics works in special relativity. The key here will be to understand the generalization of the line element concept to 4 dimensions.

In special relativity, two events $\mathcal{P}$ and $\mathcal{Q}$ separated by $\Delta x$ in space and $\Delta t$ and time can have their physical separation measured by a ruler. (I'll put the separation in the $x$-direction for simplicity.) This rule – in its own rest frame – has one end at $\mathcal{P}$ and the other *simultaneously* at $\mathcal{Q}$. If $\Delta t = t_{\mathcal{Q}} - t_{\mathcal{P}} \neq 0$, then the ruler must be moving in order for this to occur. Let's suppose that the ruler's velocity with respect to the lab frame is $\beta_x$ (in the $x$-direction) and as usual we set $\gamma = (1-\beta_x^2)^{-1/2}$. Then in the ruler's frame (which we denote with primes) the temporal and spatial separations are given by the Lorentz transformation:

$$\Delta t' = t'_{\mathcal{Q}} - t'_{\mathcal{P}} = \gamma(\Delta t - \beta_x \Delta x) \quad \text{and} \quad \Delta x' = x'_{\mathcal{Q}} - x'_{\mathcal{P}} = \gamma(\Delta x - \beta_x \Delta t). \tag{14}$$

Now if $\mathcal{P}$ and $\mathcal{Q}$ are simultaneous in the ruler frame, we need $\Delta t' = 0$ so $\beta_x = \Delta t/\Delta x$. Therefore the separation of the two points is

$$\Delta x' = (1-\beta_x^2)^{-1/2}(\Delta x - \beta_x \Delta t) = \left(1 - \frac{\Delta t^2}{\Delta x^2}\right)^{-1/2} \left(\Delta x - \frac{\Delta t}{\Delta x}\Delta t\right) = \sqrt{\Delta x^2 - \Delta t^2}. \tag{15}$$

The general spatial separation of two events in special relativity can then be written as

$$\Delta s_4^2 = \Delta s_3^2 - \Delta t^2 = \Delta x^2 + \Delta y^2 + \Delta z^2 - \Delta t^2, \tag{16}$$

where $\Delta s_4$ is the 4D separation (defined by a ruler in the frame where the two events are simultaneous), $\Delta s_3$ is the spatial separation in the lab frame, and $\Delta t$ is the temporal separation in the lab frame. If $\Delta s_4^2 < 0$, then by similar arguments we can interpret $|\Delta s_4|$ as the proper time interval separating the two events.

The above description applies to events with *finite* separation. If we want to consider the notion of a line element in curved spacetime, we should focus on infinitesimal separations. Thus in special relativity, the description of spacetime geometry is via a 4-dimensional line element that is related to the 3-dimensional line element of space as well as the time interval:

$$ds_4^2 = ds_3^2 - dt^2 = dx^2 + dy^2 + dz^2 - dt^2. \tag{17}$$

The proper time experienced by an observer traveling along some path is given as usual by

$$\tau = \int \sqrt{-ds_4^2} = \int \sqrt{1 - \left(\frac{dx}{dt}\right)^2 - \left(\frac{dy}{dt}\right)^2 - \left(\frac{dz}{dt}\right)^2}\, dt. \tag{18}$$

Paths with $ds_4^2 < 0$ everywhere are said to be **timelike** and are legal paths for matter particles to traverse. Light traverses paths with $ds_4^2 = 0$ ("**lightlike**"). Finally, paths with $ds_4^2 > 0$ are **spacelike** and cannot be traversed by any known particle.

### B. The Friedmann-Lemaître-Robertson-Walker metric

We are now ready to write down the line element for the Universe! To begin, let us imagine a model in which the spatial structure expands or contracts as a function of time, with all distances scaling in proportion to the **scale factor** $a(t)$. Then the 3-dimensional line element is

$$ds_3^2 = [a(t)]^2 \left[ \frac{dr^2}{1 - Kr^2} + r^2(d\theta^2 + \sin^2\theta\, d\phi^2) \right], \tag{19}$$

where $K = 0$ for a flat model, $K > 0$ for a closed model, and $K < 0$ for an open model. Incorporating the time dimension as we did in special relativity leads to:

$$ds_4^2 = [a(t)]^2 \left\{ d\chi^2 + [r(\chi)]^2(d\theta^2 + \sin^2\theta\, d\phi^2) \right\} - dt^2, \tag{20}$$

with $r(\chi) = \chi$, $R\sin(\chi/R)$, or $S\sinh(\chi/S)$ according to whether the Universe is flat, closed, or open. This is the line element for the **Friedmann-Lemaître-Robertson-Walker (FLRW) spacetime**, which is used to describe cosmology. It encodes all geometrical information about the expansion of the Universe, how light propagates as the Universe expands, etc.

The physical distance from us (at the origin) to an object at time $t$ is $a(t) \cdot \chi$. However, you should note that this is the distance obtained by integrating $\int ds_4$ along a path of constant $t$, $\theta$, and $\phi$; physically it is the distance that you would measure by adding up the lengths of lots of infinitesimal rulers. Since we don't measure distances in cosmology that way, none of our observables really corresponds to $a(t) \cdot \chi$ except for very nearby objects (where $\chi$ is small compared to both the age of the Universe and the radius of curvature).

An observer at fixed spatial coordinates $(r, \theta, \phi)$ is called a **comoving observer**. A comoving observer in the FLRW spacetime is non-accelerating, and measures a proper time given by the coordinate $t$.

### C. Expansion history and time variables

Distances that exclude the factor of $a(t)$ are called denoted as "comoving" and are independent of time. Since the multiplicative factor in front of $a(t)$ is arbitrary, it is common to set $a(t) = 1$ today. Since the Universe is expanding, $da/dt > 0$; in realistic cosmological models, at some time in the distant past, we would have had $a(t) = 0$. This event is known as the **Big Bang**, and corresponds to all matter in the Universe having been at a single point – an initial singularity. We normally set $t = 0$ at the Big Bang, although this is a choice of time coordinate, not a necessity. The value of time today is $t_0$.

We introduce the **Hubble rate** given by

$$H(t) = \frac{d}{dt}\ln a(t) = \frac{da/dt}{a} = \frac{\dot{a}}{a}, \tag{21}$$

and its value today, the **Hubble constant** $H_0 = \dot{a}(t = 1)$. The Hubble rate has units of 1/time. Since the separation of two galaxies scales in proportion to $a(t)$, the Hubble rate is equal to the rate of increase of separation between two galaxies ($v$), divided by their separation $s$. Rearranging this statement, we find that

$$v = Hs. \tag{22}$$

This is **Hubble's law**: the velocity with which a galaxy appears to recede ($v$) is proportional to its distance ($s$), with a coefficient of proportionality $H$. Since the above statement is instantaneous in time, it makes sense from an observational perspective only locally (i.e. at distances small enough that $H$ hasn't changed significantly in the time it takes light to get from another galaxy to us). We will explore generalizations of this idea later.

The expansion of the Universe not only stretches the separation between galaxies, but also the wavelength of light – an effect known as the **cosmological redshift**. That is,

$$\frac{\lambda_{\rm new}}{\lambda_{\rm old}} = \frac{a_{\rm new}}{a_{\rm old}}. \tag{23}$$

If we look at a distant galaxy, that means that the wavelength of light observed today (at $a_{\rm new} = 1$) is different from the wavelength at the time the light was emitted, $a_{\rm old} = a(t_{\rm em})$, in accordance with

$$\frac{\lambda_{\rm obs}}{\lambda_{\rm em}} = \frac{1}{a(t_{\rm em})}. \tag{24}$$

If the spectral line of an atom, molecule, or ion is observed from a distant galaxy, and its intrinsic wavelength is known from laboratory experiments, then $\lambda_{\rm obs}/\lambda_{\rm em}$ is directly accessible. We thus define the **redshift**

$$z = \frac{\lambda_{\rm obs}}{\lambda_{\rm em}} - 1 = \frac{1}{a(t_{\rm em})} - 1; \tag{25}$$

this is the easiest quantity to measure for a distant galaxy, far easier than measuring $t$ (the proper time elapsed from the Big Bang to the time at which the light observed today left that galaxy). For that reason, $z$ is used as the proxy for time in most of observational cosmology. It is related to the scale factor by

$$a(t_{\rm em}) = \frac{1}{1+z}, \tag{26}$$

and you will see the factor $1+z$ many times in this course. The mapping of $z$ or $a$ onto $t$ depends on the expansion history of the Universe, and measuring it is a key goal in modern cosmology.

In order to understand the paths taken by light (our principal messenger in the Universe), or to understand causal structure, it is convenient to introduce an alternative time variable: the **conformal time**

$$\eta = \int \frac{dt}{a(t)}. \tag{27}$$

Once again, we normally choose the constant of integration so that $\eta = 0$ at the Big Bang ($t = 0$). With this variable, Eq. (20) becomes

$$ds_4^2 = [a(\eta)]^2 \left\{ d\chi^2 + [r(\chi)]^2 (d\theta^2 + \sin^2\theta \, d\phi^2) - d\eta^2 \right\}. \tag{28}$$

It is thus seen that causality considerations – i.e. which paths are timelike, lightlike, or spacelike – do not depend on the expansion history except through $\eta$. In particular, light rays moving radially toward or away from the origin must have $d\chi/d\eta = \pm 1$ (the sign indicating direction of propagation).