Astrophysics and cosmology notes

Jacopo Tissino, Giorgio Mentasti, Eleonora Vanzan

2020-09-27

Contents

Introduction and relevant material

These are the (yet to be) revised notes for the course "Fundamentals of Astrophysics and Cosmology" held by professor Sabino Matarrese in fall 2019 at the university of Padua.

They are based on the notes I took during lectures, complemented with notes from the previous years.

They will be revised by the professor in the future, as of yet they have not.

The exam is a traditional oral exam, there are fixed dates but they do not matter: on an individual basis we should write an email to the professor to set a date and time.

Material There is a dropbox folder with notes by a student from the previous years [Pac18b] and handwritten notes by the professor.

There are many good textbooks, for example "*Cosmology*" by Lucchin and Coles [LC02].

Chapter 1

Cosmography

1.1 The cosmological principle

The basis for the modern treatment of cosmology is the **Copernican principle**: roughly stated, it is "*we do not occupy a special, atypical position in the universe*". We will discuss the validity of this later in this section. It is extremely useful to make such an assumption since it endows our model of the universe with a great deal of symmetry, which makes its mathematical treatment manageable.

As we will discuss shortly, this principle can be combined with our observations of isotropy to yield:

Proposition 1.1.1 (Cosmological principle)**.** *Every* comoving observer *observes the Universe around them at a* fixed time *in their reference frame as being* homogeneous *and* isotropic.

Comoving means moving coherently with the absolute reference frame, which is defined as the rest frame of the cosmic fluid, which determines the geometry of the universe.

When we observe the Cosmic Microwave Background (CMB) we see that we are surrounded by radiation distributed like a blackbody of temperature $T_{\text{CMB}} \approx 2.725\,\text{K}$ [Fix09].

This radiation is not uniformly distributed in the sky: we see a *dipole modulation* of around $\Delta T \approx 3.4\,\text{mK}$ [Col18a]. This is due to the Doppler effect: the Solar System is *not* comoving with respect to the CMB.

In fact, we can measure the *peculiar velocity* of the Solar System this way: it comes out to be around $c\Delta T / T_{\text{CMB}} \approx 370\,\text{km/s}$. This cannot be explained by the movement of the Sun through the galaxy, nor by the movement of the galaxy through the Local Group: the Local group is actually moving with respect to the absolute reference. In fact, the velocities of the Sun with respect to the Local Group and the velocity of the Local Group with respect to the CMB are almost directed in opposite directions, so the velocity of the LG with respect to the CMB can be measured to be $\approx 620\,\text{km/s}$ [Col18a, Table 3].

So, the absolute reference frame can be experimentally defined as the frame of the observer who sees the CMB with zero dipole moment.

The CMB has anisotropies of the order of $20\,\mu\text{K}$ (root mean square) [Wri03] at higher order multipole moments: it is uniform to approximately 1 part in 10^5.

The word "time" in **fixed time** refers to the proper time of a comoving observer, which is called *cosmic time*.

Homogeneity means that the characteristics of the universe as observed from a point are the same as they would be as observed from any other point.

Isotropy means that the characteristics of the universe as observed in a certain direction are the same as they would be if they were observed in another direction.

On the validity of the cosmological principle The principle is expected to hold only on very large[1] scales: at small scales we see structures, such as galaxies or our Solar System, so we surely do not have homogeneity.

Change citation for homogeneity length scale?

How can we talk about homogeneity if we can only look at the universe from a single point? We must *assume* that any other observer would also see isotropy as we do: this is precisely what the Copernican principle tells us.

Isotropy around every point implies homogeneity. We observe isotropy, and with the assumption that we are typical observers we obtain homogeneity.

In the end, this assumption is the basis of modern cosmology: it has to be made before any cosmological study starts: it might not be completely correct, but it allows us to make falsifiable predictions, so we shall keep it until the models it allows us to create do not match observations anymore.

1.2 The geometry of spacetime

The best description for gravity we have so far is given by the general theory of relativity (GR). In it, spacetime is modelled as a 3+1-dimensional semi-Riemannian manifold with a line element which is generally given by the expression $\mathrm{d}s^2 = g_{ab}\,\mathrm{d}x^a\,\mathrm{d}x^b$ and which.

Latin indices can take values from 0 to 3, and we adopt the "mostly minus" metric signature.

Such a spacetime can have up to $4(4+1)/2 = 10$ global continuous symmetries, which can be classified into:

(a) 1 time translation;

(b) 3 Lorentz boosts;

(c) 3 spatial translations;

(d) 3 spatial rotations.

The metric of Minkowski spacetime, which has all of these 10 symmetries, reads:

$$\mathrm{d}s^2 = c^2\,\mathrm{d}t^2 - \mathrm{d}\vec{x}^2 \qquad \text{(Cartesian coordinates)} \tag{1.1a}$$

[1] Larger than $260h^{-1}$Mpc ≈ 380 Mpc $\approx 1.2 \times 10^{25}$ m [YBK10], where $h \approx 0.68$ [Col16]: this will be made clearer in later sections, but is meant to give an idea of the length scales involved . The portion of Universe we can see is of order 10 Gpc.

$$= c^2\,\mathrm{d}t^2 - \mathrm{d}r^2 - r^2\,\mathrm{d}\Omega^2 \qquad \text{(spherical coordinates)}\,, \tag{1.1b}$$

where $r = |x|$, while θ and φ are the spherical angles and $\mathrm{d}\Omega^2 = \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2$.
Minkowski space is *maximally symmetric*, which means we could not have more symmetries than these. It is not the only possibility, there are three maximally symmetric spacetimes: Minkowski, de Sitter, anti-de Sitter.

In cosmology we lose time translation symmetry, since the universe is expanding, and Lorentz boost symmetry, since as we saw in the previous section we can tell at which speed we are moving with respect to the CMB.

We keep the purely spatial symmetries: so, our description of spacetime will be as a 3+1-dimensional manifold which, if we fix the temporal coordinate for a comoving observer, reduces to a *maximally symmetric* 3-dimensional space — for 3 dimensions the maximum number of symmetries is $3(3+1)/2 = 6$, which correspond to (c) and (d).

It can be shown that the most general form of a metric satisfying these conditions is the **Friedmann-Lemaître-Robertson-Walker line element**,[2] which, in the comoving frame, reads

$$\boxed{\mathrm{d}s^2 = c^2\,\mathrm{d}t^2 - a^2(t)\left(\frac{\mathrm{d}r^2}{1-kr^2} + r^2\,\mathrm{d}\Omega^2\right).} \tag{1.2}$$

The coordinate r does not have the dimensions of a length: we choose our variables so that $a(t)$ is a length, while r is dimensionless.[3]

The parameter $a(t)$ is called the *scale factor*: varying it amounts to rescaling all of space. It has the dimensions of a length, and it depends on the cosmic time t.

The parameter k is a constant describing the spatial curvature, which can always be normalized to ± 1 or 0. Universes with:

- $k = 1$ are called closed universes;
- $k = -1$ are called open universes;
- $k = 0$ are called flat universes.

We can choose the normalization of k, but its sign is a constant. Positive values correspond to $k = 1$, negative values to $k = -1$. When computing probability distributions for k, we must not consider it to be a discrete variable but a continuous variable instead: so, the set of flat universes with $k = 0$ has zero measure, and thus zero probability with any probability density function.

[2] This is sometimes called just "Robertson-Walker", or RW, or FLRW.

[3] This is a matter of convention: if we normalize $|k| = 1$ then we will have $0 \le r < 1$; we could also let k be arbitrarily large; then r would also be.

1.2.1 A bidimensional example

We consider surfaces: these are the simplest manifolds which can have intrinsic curvature.

Intrinsic curvature is described by the Riemann tensor, which has 20 independent components in 4D and 6 in 3D, so it is difficult to visualize.

On the other hand, in 2D has only 1 independent component: $R_{1212} = R \det g_{ab}$, where R is the scalar curvature.

The scalar curvature R has an immediate geometric interpretation: it is equal to $2/(r_1 r_2)$, where r_i are the radii of the osculating circles at the point; it is positive if the circles are in the same direction, as they are for a sphere, and negative if they are in different directions, as they are for a hyperboloid. For a flat surface we cannot define an osculating circle (at least not in both directions): its radius diverges, so the curvature vanishes.

The metric for a Cartesian flat 2D plane is $\mathrm{d}l^2 = a^2(\mathrm{d}r^2 + r^2\,\mathrm{d}\theta^2)$. The constant a is included since we want r to be dimensionless.

The metric for the surface of a sphere is: $\mathrm{d}l^2 = a^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right)$, where $a^2 = R^2$, the square radius of the sphere.

The metric for the surface of a hyperboloid is: $\mathrm{d}l^2 = a^2\left(\mathrm{d}\theta^2 + \sinh^2\theta\,\mathrm{d}\varphi^2\right)$, therefore the only difference is that trigonometric functions become hyperbolic ones.

Do note that for the sphere $\theta \in [-\pi, \pi]$ while for the hyperboloid we can in principle have $\theta \in \mathbb{R}$: this is indicative of the fact that the sphere is bounded, while the hyperboloid is not.

For both of these, let us define the variable: $r = \sin\theta$ in the spherical case, and $r = \sinh\theta$ in the hyperbolic case.

As we change variable we do the following manipulation for the sphere:

$$\mathrm{d}r^2 = \left(\frac{\mathrm{d}r}{\mathrm{d}\theta}\right)^2 \mathrm{d}\theta^2 = \cos^2\theta\,\mathrm{d}\theta^2 \tag{1.3a}$$

$$\implies \mathrm{d}\theta^2 = \frac{\mathrm{d}r^2}{\cos^2\theta} = \frac{\mathrm{d}r^2}{1-r^2}\,, \tag{1.3b}$$

and similarly for the hyperboloid, except for the fact that in that case $\cosh^2\theta = 1 + \sinh^2\theta = 1 + r^2$.

So, the line elements become respectively:

$$\mathrm{d}l^2{}_{\text{sphere}} = a^2\left(\frac{\mathrm{d}r^2}{1-r^2} + r^2\,\mathrm{d}\varphi^2\right) \tag{1.4a}$$

$$\mathrm{d}l^2{}_{\text{hyperboloid}} = a^2\left(\frac{\mathrm{d}r^2}{1+r^2} + r^2\,\mathrm{d}\varphi^2\right). \tag{1.4b}$$

We have a striking similarity to the Robertson-Walker metric: we only need to make the substitution $\mathrm{d}\varphi \to \mathrm{d}\Omega$ in order to recover it.

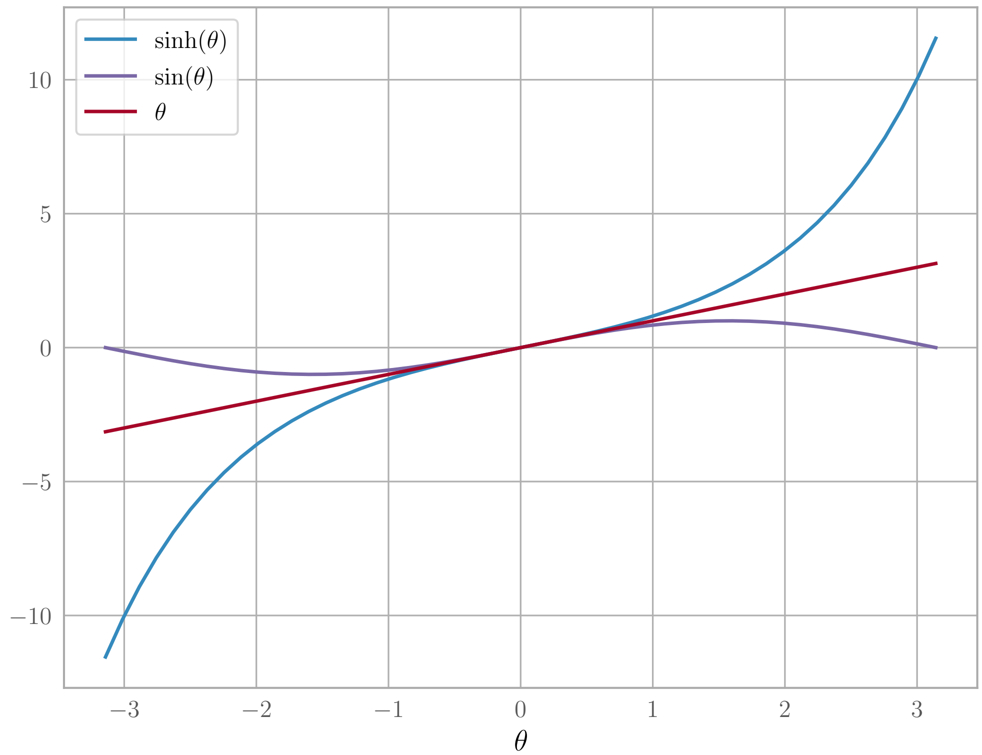

Figure 1.1: Plot of the functions $\sin(\theta)$ and $\sinh(\theta)$ in the interval $-\pi \leq \theta \leq \pi$.

We can also work backwards and rewrite the RW line element in sphere- or hyperboloid-like coordinates:

$$\mathrm{d}l^2 = c^2\,\mathrm{d}t^2 - a^2 \begin{cases} \mathrm{d}\chi^2 + \sin^2\chi\,\mathrm{d}\Omega^2 \\ \mathrm{d}\chi^2 + \chi^2\,\mathrm{d}\Omega^2 \\ \mathrm{d}\chi^2 + \sinh^2\chi\,\mathrm{d}\Omega^2 \end{cases} \tag{1.5}$$

where we introduce a variable χ defined so that if $k = +1$ then $r = \sin\chi$, if $k = 0$ then $r = \chi$, and if $k = -1$ then $r = \sinh\chi$.

The properties of the sphere and of the hyperboloid actually carry over to the 3D case: a spacetime with positive curvature is bounded, while if it is flat or hyperbolic it is unbounded.

1.2.2 Other forms of the RW metric

If we wish to use Cartesian coordinates the RW metric takes the following expression:

$$\mathrm{d}s^2 = c^2\,\mathrm{d}t^2 - a^2(t)\left(1 + \frac{k|x|^2}{4}\right)^{-2}\left(\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2\right). \tag{1.6}$$

Universes in which a is a constant are called *Einstein spaces*.

We can also change time variable: the *conformal time* η is such that $\mathrm{d}t = a(\eta)\,\mathrm{d}\eta$, where $a(\eta) \stackrel{\text{def}}{=} a(t(\eta))$: so, we will have

$$\mathrm{d}s^2 = a^2(\eta)\left(c^2\,\mathrm{d}\eta^2 - \left(\frac{\mathrm{d}r^2}{1-kr^2} + r^2\,\mathrm{d}\Omega\right)\right). \tag{1.7}$$

Clarify distinction on different types of conformal transformations and dependence on spatial curvature.

Weyl transformations are defined to be those which preserve angles locally; since angles are defined through the metric but the angle between two vectors does not change if they are rescaled, this can be translated into the condition that the metric is rescaled by a generic function:

$$g_{ab} \to a^2(x^i) g_{ab}\,. \tag{1.8}$$

If two metric are mapped into each other by a Weyl transformation, they are said to be in the same *conformal class*.

In our case, the dependence on the point in spacetime is reduced to a dependence on the cosmic time only since we have symmetry with respect to spatial translations.

The conformal time is called that because if we use it we can map our spacetime at any time to spacetime at another time using a Weyl transformation. If there is no spatial curvature, we can also map it to flat Minkowski spacetime.

The universe we inhabit does not have conformal symmetry: a generic massive particle in it has a Compton wavelength $\lambda = h/(mc)$ which defines its interaction cross section, so if the spacetime expands an ensemble of these particles will have different dynamics.

However, conformal geometry is useful for the description of particles which have no characteristic length, such as photons. Particles with no characteristic length are insensitive to dilation, since they do not have a "meter" to probe the expansion of spacetime. The photons of the CMB look like they are thermal: they were thermal in the early universe, since they were in thermal equilibrium with matter (photons and matter particles were constantly Compton-scattering off each other); when matter and radiation decoupled the photons scattered for the last time and then kept travelling. The universe has since expanded by a factor ≈ 1090, but because of the fact that the photons do not have a characteristic length their distribution can still be modelled as a blackbody distribution for an appropriately rescaled temperature. However, strictly speaking, we are not allowed to say that they are thermal, since keeping thermal equilibrium implies that interactions are occurring.

Notice that we have made no use of dynamics so far: we wrote the line element, the solution of the Einstein equations, without the Einstein equations themselves! Of course we can obtain the Robertson-Walker metric starting from the field equations as well, but here we have only based our considerations on geometrical assumptions. This approach is called **Cosmography**.

1.3 The energy budget of the universe

Up until now we did not consider any dynamics in our spacetime. We will discuss this topic in more detail in later sections, but for now we give the result: the dynamics of the universe are described by the Friedmann equations:

$$\dot{a}^2 = \frac{8\pi G_N}{3}\rho a^2 - kc^2 \tag{1.9a}$$

$$\ddot{a} = -\frac{4\pi G_N}{3}a\left(\rho + \frac{3P}{c^2}\right) \tag{1.9b}$$

$$\dot{\rho} = -3\frac{\dot{a}}{a}\left(\rho + \frac{P}{c^2}\right) \tag{1.9c}$$

where dots denote differentiation with respect to the *cosmic time* t, G_N is Newton's gravitational constant, $\rho = \rho(t)$ is the energy density, and $P = P(t)$ is the isotropic pressure.

The curvature k appears in the first equation: so we can try to measure it by comparing the other two terms in the equations. This way, we can determine whether the universe is flat or curved.

In order to discuss this problem, let us establish some notation: an important parameter is $H(t) \stackrel{\text{def}}{=} \dot{a}/a$, the *Hubble parameter*. We can write an equation for it from the first Friedmann one:

$$H^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2} \tag{1.10}$$

If $k = 0$, then there we must have a critical energy density

$$\rho_C(t) = 3H^2(t)/(8\pi G)\,, \tag{1.11}$$

and we define $\Omega(t) = \rho(t)/\rho_C(t)$. For a flat universe, $\Omega = 1$, and so we can determine $k = \operatorname{sign}(\Omega - 1)$, since:

$$1 = \frac{8\pi G}{3}\frac{\rho}{H^2} - k\frac{c^2}{a^2H^2} \tag{1.12}$$

Dividing equation (1.10) through by H^2

$$\frac{\rho}{\rho_c} - 1 = k\frac{c^2}{a^2H^2} \tag{1.13}$$

$$\operatorname{sign}(\Omega - 1) = \operatorname{sign}\left(k\frac{c^2}{a^2H^2}\right) = \operatorname{sign}(k) = k\,. \tag{1.14}$$

Took sign of both sides

This is a promising way to measure the curvature of the universe. As we will see, we can infer the densities of the various constituents of the universe through their dynamics. Notice that the measurement of the energy density is a "Newtonian" measurement, while that of the geometry of spacetime is a General Relativity one.
The alternative is trying to measure the curvature geometrically, however we should look at really large scales in order to see any effects: we are thus drawn to the CMB. Unfortunately, the geometrical effects of spatial curvature on the CMB power spectrum[4] are highest for

[4] The power spectrum is, roughly speaking, the set of the square moduli of the coefficients in the spherical-harmonics decomposition, classified according to the coefficient l; the higher l, the smaller the angular scale we are considering.

the lowest multipoles, for which we have the largest variance. This means that the direct geometric effects of curvature cannot be discerned in the CMB power spectrum, however the dynamical effects can.

Further, we define the **Hubble constant**:

$$H_0 = H(t_0) = 100h \times \mathrm{km\,s^{-1}\,Mpc^{-1}} \approx 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}} \tag{1.15}$$

where $h \approx 0.7$ is a number, and t_0 just means *now*.[5]

The reason for this peculiar way to write the constant is that historically it has been difficult to determine the value of H_0 precisely, and it affects many astronomical conversions: keeping it indeterminate in this way allows us to quickly update our old estimates if we measure H_0 more precisely later. Historically, in the American school, the pupils of Hubble thought $h \sim 0.5$, while the French school thought $h \sim 1$. Now, a great issue in cosmology is the disagreement between the measurements of H_0 obtained from the cosmic distance ladder and those obtained from the CMB [Won+19].

1.3.1 Energy density

In order to determine the spatial curvature of the universe we need to look at $\Omega \propto \rho$: we need to measure the energy density of the universe.

How do we do it?

Let us start by considering the energy density *today* (index 0) due to *galaxies* (index g): ρ_{0g}. We do not directly observe the mass[6] of galaxies: we can only measure their luminosities.

So, we do the following: we compute the mean value of ρ, the mass per unit volume, with the aid of the galaxy luminosity per unit volume ℓ: the mean density is given by the mean luminosity times the average ratio of mass to luminosity of galaxies:[7]

$$\langle \rho \rangle \sim \langle \ell \rangle \left\langle \frac{M}{L} \right\rangle , \tag{1.18}$$

where $\langle M/L \rangle$ is the average ratio of mass over luminosity per galaxy: we had a ratio of densities, but since we are considering averages we can integrate above and below with respect to the spatial volume.

[5] Also, pc means "parsec":

$$1\,\mathrm{pc} = \frac{1\,\mathrm{AU}}{1\,\mathrm{arcsec}} \approx 3.085 \times 10^{16}\,\mathrm{m} \approx 3.26\,\mathrm{ly}\,, \tag{1.16}$$

where the angle is to be interpreted as dimensionless (in radians), and AU is an astronomical unit, the Earth-Sun average distance. The definition is as such because of a way we have to measure the distances to nearby objects, by measuring their parallax between winter and summer: if they are close enough, they will have apparently moved with respect to further objects, because of the movement of the Earth around the Sun.

[6] Here the terms mass and energy are used equivalently: the velocity of galaxies with respect to the CMB is nonrelativistic, so we approximate $E = \gamma mc^2 \approx mc^2$.

[7] Formally, the steps are

$$\langle \rho \rangle = \left\langle \ell \, \frac{\rho}{\ell} \right\rangle \approx \langle \ell \rangle \left\langle \frac{\rho}{\ell} \right\rangle = \langle \ell \rangle \left\langle \frac{M}{L} \right\rangle , \tag{1.17}$$

where we made the assumption of the ratio ρ/ℓ being *uncorrelated* to the luminosity ℓ. This is not precisely verified, but we are giving order-of-magnitude estimates so this is close enough for our purposes.

It is measured in units of solar mass over solar luminosity: $M_\odot/L_\odot$. Reference values for these are $M_\odot \sim 1.99 \times 10^{33}\,\mathrm{g}$, while $L_\odot \sim 3.9 \times 10^{33}\,\mathrm{erg\,s^{-1}}$.

We denote $\langle \ell \rangle \stackrel{\text{def}}{=} \mathscr{L}_g$: it is the mean (intrinsic, bolometric[8]) luminosity of galaxies per unit volume.

By definition, it is given by

$$\mathscr{L}_g = \int_0^\infty L\Phi(L)\,\mathrm{d}L \,, \tag{1.19}$$

where $\Phi(L)$ is the number density of galaxies per unit volume and unit luminosity: the *luminosity function*.

The Schechter function is an empirical estimate for the shape of this distribution:

$$\Phi(L) = \frac{\Phi^*}{L^*}\left(\frac{L}{L^*}\right)^{-\alpha} \exp\left(-\frac{L}{L^*}\right), \tag{1.20}$$

where Φ_*, L_* and α are parameters, with dimensions of respectively a number density, a luminosity and a pure number.

These can be fit by observation: we find $\Phi^* \approx 10^{-2}h^3\mathrm{Mpc}^{-3}$, $L^* \approx 10^{10}h^{-2}L_\odot$ (a typical galaxy contains roughly ten billion Suns) and $\alpha \approx 1$.

The integral for $\mathscr{L}_g$ converges despite the divergence of $\Phi(L)$ as $L \to 0$, since it is multiplied by L: so we do not need to really worry about the low-luminosity divergence of the distribution.

The result of the integral for a generic value of α is $\mathscr{L}_g = \Phi^* L^* \Gamma(2-\alpha)$, where Γ is the Euler gamma function; for the $\alpha = 1$ case we get a factor $\Gamma(2-1) = 1$.

Numerically, inserting reasonable estimates for the parameters, we get the following estimate for the mean luminosity: $\mathscr{L}_g \approx (2.0 \pm 0.7) \times 10^8 h L_\odot \mathrm{Mpc}^{-3}$.

Now, we must estimate $\langle M/L \rangle$. The luminosity of galaxies can be measured readily, the great difficulty lies in estimating their mass.

1.3.2 Estimating the masses of galaxies

We must distinguish between the different shapes of the galaxies: spiral galaxies are characterized by rotation of the stars about the galactic center, while in elliptical galaxies the stars' motion is disordered.

Spiral galaxies

If we see a spiral galaxy edge-on, we will have a side of it coming in our direction, and the other side moving away from us (after correcting for other sources of Doppler shift, such as the velocity of the whole galaxy). So, using the Doppler effect we can measure the distribution of the velocity in the galaxy as a function of the radius.

In order to get a theoretical model, we can approximate the galaxy as a sphere: this is very rough (spiral galaxies are closer to being disk-like), but it gives the same qualitative result, so there is no need for a more precise model in this context.

[8] Bolometric means "total, over all wavelengths", as opposed to the luminosity in a certain wavelength band, which is easier to measure in astronomy.

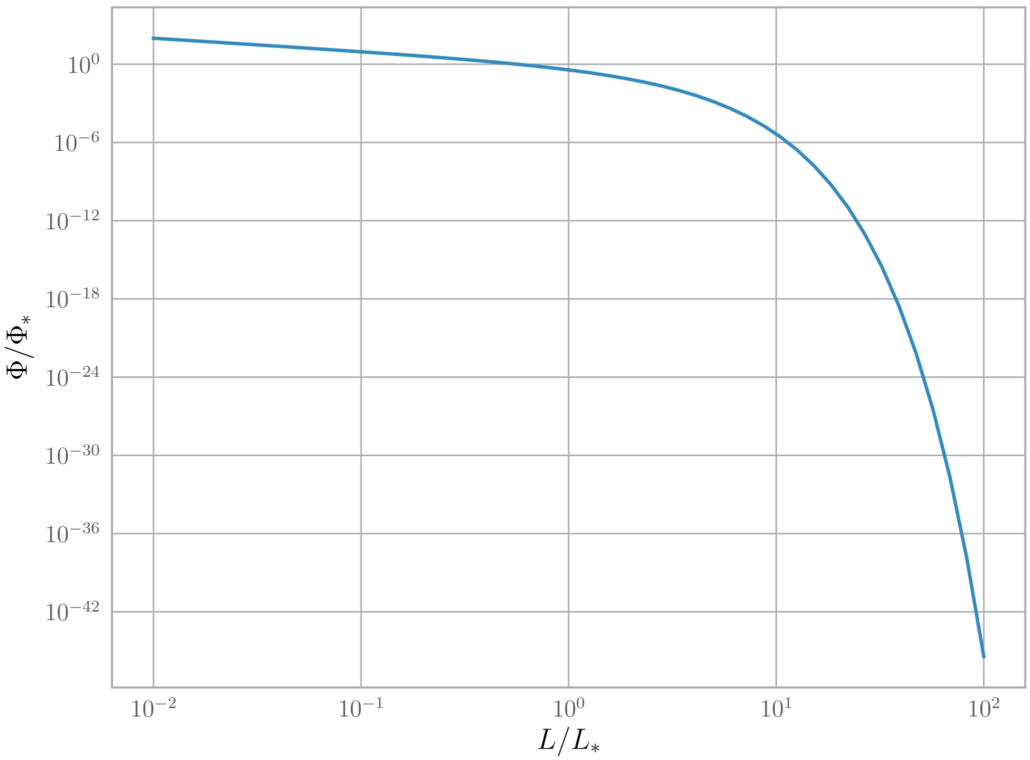

Figure 1.2: A rough plot of the Schechter function for $\alpha = 1$.

We model the galaxy velocity distribution using Newtonian mechanics: the GR corrections are negligible at these scales, galaxies are much larger than their Schwarzschild radii.

Equating the gravitational acceleration GM/R^2 to the centripetal acceleration v^2/R we find:

$$v = \sqrt{\frac{GM}{R}}\,, \tag{1.21}$$

where v and M are functions of R: M is the mass contained in the spherical shell of radius R, and v is the orbital velocity at the boundary of the shell.

In the inside regions of the galaxy, where $M(R) \propto R^3$ since the density is approximately constant, we will have $v \propto R$, while in the outskirts of the galaxy $M(R)$ will not change much, since all the mass is inside, so we will have $v \propto R^{-1/2}$.

Our prediction is then a roughly linear region, and then a region with $v \sim R^{-1/2}$. This is shown as the "Predicted" curve in figure 1.3.

Instead of this, when in the 1980s people started to be able to measure this curve accurately they saw that, after the linear region, $v(R)$ was approximately constant. So, is Newtonian gravity wrong?

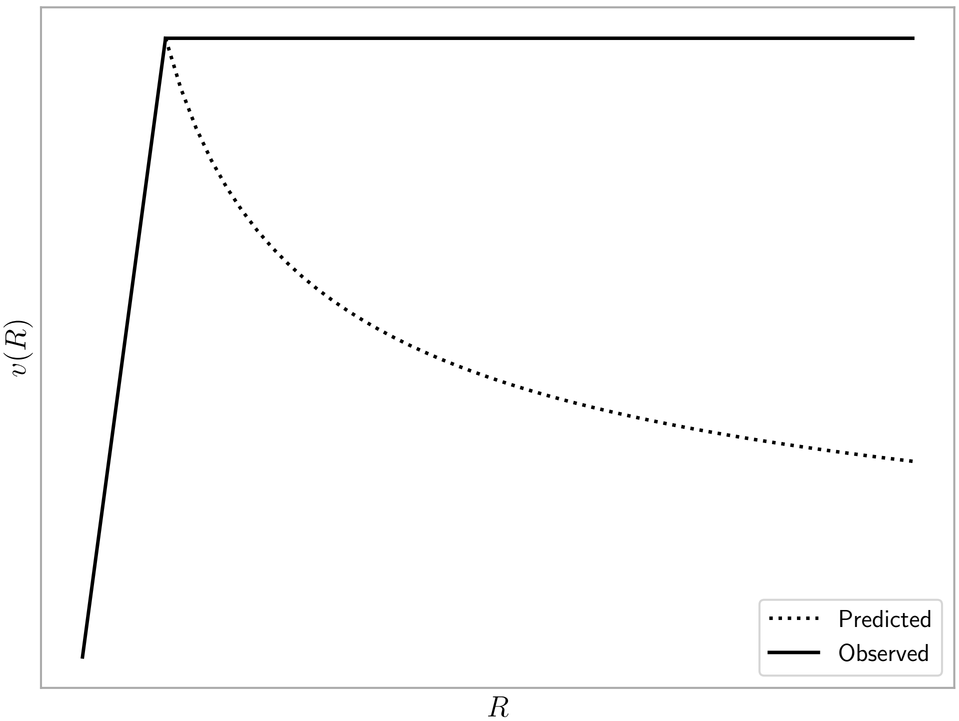

Figure 1.3: Predicted and observed velocity distribution for galaxies. The point at which they star to diverge is approximately the radius of the bulk of the galaxy. This plot is approximate, realistic models do not have such sharp corners, since there will not be a precise edge after which the density drops to zero immediately.

An option to solve this problem was proposed by Milgrom and collaborators: it is called MOdified Newtonian Dynamics, or MOND: they propose that gravity is not actually always described by a r^{-1} potential, but instead at low accelerations it behaves differently. Specifically, they posit that the gravitational acceleration g should be modulated by a factor $\mu(g/a_0)$, where $a_0 \approx 8 \times 10^{-10} h^2 \mathrm{m/s}^2$ is a characteristic acceleration while $\mu(x)$ is an dimensionless function such that $\mu \to 1$ for $x \gg 1$ and $\mu \to x$ for $x \ll 1$, such as $\mu = x/(1+x)$ [BM84].

This approach is Newtonian but there are also relativistic MOND variants. They do not match observation as well as the alternative approach does.[9] The heaviest thing weighing against MOND is the fact that even using it we still need dark matter in order to fully explain observations.

[9] MOND would be compatible with the speed of gravity being less than the speed of light, which is equivalent to the graviton being massive. Recent measurements of gravitational waves seem to agree with the general-relativistic prediction that it is massless.

Dark matter

The alternative option is that Newtonian mechanics describes galactic mechanics well, but the galaxy's matter distribution inferred from our observations is actually smaller than the real distribution, which extends outward further than the matter we can see: this is **dark matter**.

We'd need mass obeying $M(R) \propto R$ in order to have a constant value of v: since $M(R) = 4\pi \int_0^{R_{\max}} R^2 \rho(R)\, \mathrm{d}R$, we need the density profile to decay like $\rho(R) \propto R^{-2}$. This is called an *isothermal* density profile: we call it the *dark matter halo*, which surrounds all spiral galaxies.

We do not know what dark matter is: we can say that it interacts gravitationally but not electromagnetically. People tend to believe that it is made up of beyond-the-standard-model particles, like a *neutralino* or an *axion*. Historically, people thought the effect could be due to massive neutrinos; however their mass would need to be around 30 eV, and the analysis of the CMB data showed that the sum of the masses of the neutrino species must be $\sum m_\nu < 0.120\,\mathrm{eV}$ [Col18a, Table 7].

The total density of matter (dark+regular) is ~ 6 times more than that of regular matter alone: this must be accounted for in our estimate of the M/L ratio for spiral galaxies (since we have additional mass but not additional luminosity): with this correction, we find that for spiral galaxies

$$\left\langle \frac{M}{L} \right\rangle \approx 300h \frac{M_\odot}{L_\odot}\,, \tag{1.22}$$

Historically, this was the first evidence for dark matter.

Elliptical galaxies

If galaxies are not spiral-shaped, we have to weigh them in a different way: the Doppler broadening of spectral lines gives us a measure of the root-mean-square velocity.

Later in the course we will obtain the (nonrelativistic) *virial theorem*, now we just state it: if T is the kinetic energy of a gravitationally bound system, U is the potential energy, then

$$2T + U = 0 \tag{1.23}$$

holds when the inertia tensor stabilizes, that is, when we have dynamical equilibrium.

The kinetic energy is $T = \frac{3}{2} M \left\langle v_r^2 \right\rangle$, where v_r is the radial[10] component of the velocity, which we expect to account for one third of total energy by the equipartition theorem. M is the total mass of the galaxy.

The potential energy, instead, is $U = -GM^2/R$. Substituting these expressions into the virial theorem we get

$$2 \times \frac{3}{2} M \left\langle v_r^2 \right\rangle - \frac{GM^2}{R} = 0 \tag{1.24}$$

[10] By "radial" we mean directed towards us, not towards the center of the galaxy.

$$M = \frac{3R}{G} \left\langle v_r^2 \right\rangle , \tag{1.25}$$

so if we can measure $\left\langle v_r^2 \right\rangle$ through Doppler broadening and we can give a reasonable estimate for the radius R of the galaxy we can give an estimate for M.

Global matter contributions

Accounting for the dark matter mass, we get $\langle M/L \rangle \approx 300 h M_\odot / L_\odot$.

The value of the critical energy density today, ρ_{0c}, is given by[11]

$$\rho_{0c} = \frac{3H^2}{8\pi G} \approx 1.88 \times 10^{-29} h^2 \text{g}/\text{cm}^3 , \tag{1.26}$$

so, in order to have $\Omega = 1$, we'd need $\langle M/L \rangle$ to be equal to:

$$\left\langle \frac{M}{L} \right\rangle = \frac{\rho_{0c}}{\mathscr{L}_g} \approx \frac{1.88 \times 10^{-29}\, h^2 \text{g}/\text{cm}^3}{2 \times 10^8\, h L_\odot / \text{Mpc}^3} \approx 1390 h M_\odot / L_\odot . \tag{1.27}$$

We can define quantities of the form $\Omega_{0i} = \rho_{0i}/\rho_{0c}$, where i is a type of matter, such as baryonic matter, dark matter, dark energy, radiation and so on, whose density is represented as ρ_{0i}.

These variables quantify how much, at the present, time, of the cosmic energy budget is accounted for by that type of matter.

So, only $\Omega_{0b} \approx 5\%$ of the energy budget is given by baryonic matter (not all of which is visible), while around $\Omega_{0DM} \approx 27\%$ is dark matter. Together, they are just denoted as "matter", and $\Omega_{0m} \approx 30\%$.

We can ask ourselves: is dark matter actually baryonic matter which for some reason we cannot see, such as black holes or brown dwarfs?

This cannot be the case: our observations, combined with models for primordial nucleosynthesis, gives the following bounds for the baryonic energy density:

$$0.013 \leq \Omega_{0\text{b}} h^2 \leq 0.025 . \tag{1.28}$$

The upper bound for Ω_{0b} is around $2.5\%/h^2 \approx 5.4\,\%$.

This would seem to indicate that $\Omega_0 \approx 0.3 \ll 1$: however we are failing to consider a crucial contribution. Consider the second Friedmann equation (1.9b): in the Newtonian limit $P \sim 0$ while $\rho > 0$, so we get $\ddot{a} < 0$: the universe contracts. This is not what is observed: we actually see it in accelerated expansion.

1.3.3 Dark energy

The measurements leading to this conclusion are performed by estimating the distance and redshift of far-away objects whose intrinsic luminosity is well known, called *standard candles*: the most commonly used are type Ia supernovae and Cepheid variables.

[11] Do note that the numerical figure, 1.88×10^{-29}, is approximate but its value is known to at least four significant digits, since the only source of uncertainty in it lies in the uncertainty in our measurement of G — all the uncertainty in H_0 is expressed in variable form, with the parameter h.

So, if the expansion is accelerated then $\ddot{a} > 0$: this means, again from the second Friedmann equation (1.9b), that $P < -\rho c^2/3$. This is commonly expressed by defining $w = P/\rho c^2$. In order to have accelerated expansion we need $w < -1/3$; what is observed is closer to $w \sim -1$.

This negative pressure has the effect of a *tension*, pulling the universe apart. As we will discuss in section 2.5, a candidate for a cosmological fluid with negative pressure (specifically, $w = -1$) is a cosmological constant term, called Λ, which can be inserted in the Einstein Field Equations. It is not the possible one: dark energy is defined to be what causes the expansion we see, so it could be constituted by any kind of fluid which is uniformly distributed in space and which has negative pressure.

We cannot see directly neither dark matter nor dark energy: how do we distinguish the two? Dark matter tends to cluster, while dark energy is uniformly distributed; furthermore, dark energy has negative pressure.

From observations of both the anisotropies in the CMB and the distribution of galaxies we can determine that $\Omega_\Lambda \approx 0.7$.

1.3.4 Radiation

We still need to compute the contribution of the energy of electromagnetic and neutrino radiation to the total energy balance. Let us start with EM radiation: the greatest fraction of the radiation energy density is contained in the CMB,[12] which is extremely close to a Planckian distribution:

$$B(\nu, T) = \frac{2h}{c^2} \frac{\nu^3}{\exp\left(\frac{h\nu}{k_B T}\right) - 1}, \tag{1.29}$$

with $T = T_{0\gamma} \approx (2.725 \pm 0.001)\,\mathrm{K}$. B is a measure of spectral intensity: it is measured in units of energy per unit second, area, solid angle and frequency. This is a well-known distribution, whose integral is given by

$$\rho_{0\gamma} = \frac{\sigma_r T_{0\gamma}^4}{c^2} = 4.6 \times 10^{-34}\,\mathrm{g\,cm^{-3}} \tag{1.30}$$

[12] This fact is not obvious: the CMB permeates the universe, but each photon in it has a much lower energy than the typical photon emitted by a star. Also, stellar fusion is quite efficient, being able to convert around 0.7 % of the mass of its baryons into energy in the form of photons. If all the baryonic matter in the universe underwent fusion, this would give us a contribution of the order of $\Omega_m \times 0.7\,\% \approx 5\,\% \times 0.007 \approx 3 \times 10^{-4}$, which as we will see in a moment is of the same order of magnitude as the CMB contribution.

However, assuming that all baryonic matter undergoes fusion leads us to overestimate the result: first, a large fraction of the baryons in the universe are not in star-forming galaxies but are instead distributed in filaments among them, in what is called the Warm-Hot Intergalactic Medium, which accounts for around 30 % of the baryonic density in the universe [dGra+19].

Also, in a given galaxy only a small fraction of the baryonic mass is contained in stars: roughly speaking, 15 % [And10, fig. 11].

A third point: stars only started forming in the relatively-early universe, with redshifts of around $z \sim 10 \div 30$.

As for high-energy (UV) photons, their contribution is truly negligible: their intensity is around $3\,\mathrm{nW/m^2/sr}$ [Col18b, page 2], corresponding to a density of around $10^{-36}\,\mathrm{g/cm^3}$ (we multiply by $4\pi/c^3$ to find this result). The same paper estimates the infrared density at around 4 times the UV one, which still means it is a couple orders of magnitude below the CMB density.

where $\sigma_r = \pi^2 k_B^4/(15\hbar^3 c^3)$, while $\sigma_{SB} = \sigma_r c/4$.

So, the radiation contribution to the global energy balance is $\Omega_{0\gamma} \approx 2.5 \times 10^{-5}\, h^{-2} < 0.01\,\%$, definitely negligible.

We are going to show in later sections that if neutrinos were massless, their temperature would be $T_\nu = (4/11)^{1/3} T_\gamma < T_\gamma$.

They might not be massless, and if they are not the main contribution to the energy density they will give will be from their masses. However, recent observations (for example, by the Planck satellite) are bounding the mass of the neutrinos, $\sum m_\nu \leq 0.12\,\mathrm{eV}$: we have

$$\rho_\nu = 3N_\nu \frac{\langle m_\nu \rangle}{10\,\mathrm{eV}} 10^{-30}\,\mathrm{g\,cm^{-3}}\,, \tag{1.31}$$

where N_ν is the number of neutrino species. Even if we assume the upper bound, $N_\nu \langle m_\nu \rangle = 0.12\,\mathrm{eV}$, we get $\Omega_{0\nu} < 0.5\,\%$.

Conclusions

If we add up all the contributions to $\Omega = \Omega_b + \Omega_{DM} + \Omega_\Lambda$ (neglecting, as we said, the contributions by EM radiation and neutrinos) we find experimentally $\Omega_k \overset{\text{def}}{=} 1 - \Omega \approx (5^{+38}_{-40}) \times 10^{-4}$ [Pla+19, Table A.2].

So, with the observational uncertainties we have currently we cannot determine the sign of the universe's spatial curvature: the value of Ω_k is very much compatible with 0; even though one is drawn to say that this means that an open universe is more likely, no particular meaning should be drawn from the fact that the nominal value of Ω_k is slightly above 0.

1.3.5 The Hubble law

At the end of the 1920s Edwin Hubble compared the estimates for the distances of far-away galaxies (obtained through standard candles and other types of estimates) to their velocities relative to us, as measured through their redshift. His results are shown in figure 1.4: he obtained a roughly linear relation in the form:

$$v = H_0 d\,, \tag{1.32}$$

where v is the velocity of the galaxies, and d is their distance from us, while H_0 is a constant of proportionality. Hubble's measurements suggested $H_0 \sim 500\,\mathrm{km/s/Mpc}$, more refined modern ones using techniques similar to those used by Hubble yield a value of $(73.24 \pm 1.74)\,\mathrm{km/s/Mpc}$ [Rie+16].

Measurements of H_0 through analysis of the CMB gives an incompatible value: $H_0 = (67.8 \pm 0.9)\,\mathrm{km/s/Mpc}$ [Col16].

Let us now show that this H_0 is actually the same one we defined before, $H_0 = \dot{a}/a$. H_0 is called the **Hubble constant**, since it is constant with respect to the direction we look in the sky.

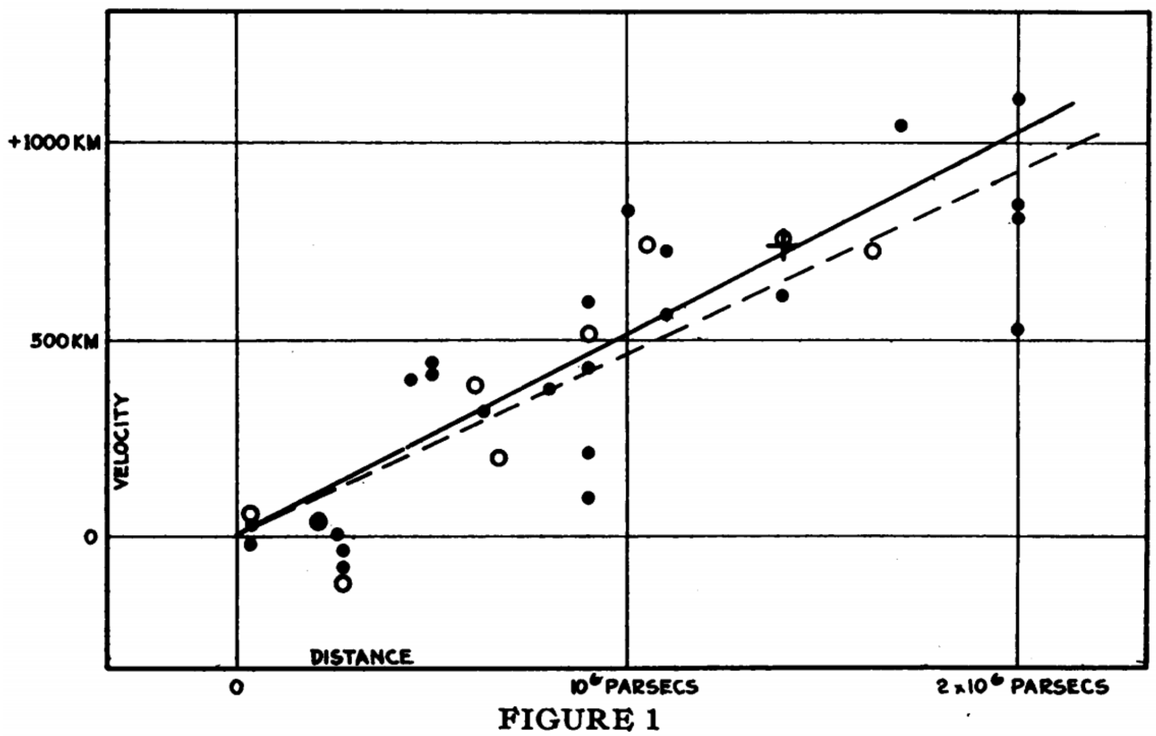

Figure 1.4: Original velocity versus distance data from Edwin Hubble's paper in 1929 [Hub29].

We are considering the distance connecting us (the center of our reference frame) to a distant galaxy, so we drop the angular part in the flat ($k = 0$) FLRW line element:

$$\mathrm{d}s^2 = c^2\,\mathrm{d}t^2 - a^2(t)\,\mathrm{d}r^2 \ . \tag{1.33}$$

So, at a fixed time the physical distance is given by $d = a(t)r$: therefore $v = \dot{d} = \dot{a}r = \frac{\dot{a}}{a}d = H_0 d$.

Assuming the universe is spatially flat is correct up to second order: for a general value of k we have

$$d = a\int_0^r \frac{\mathrm{d}\widetilde{r}}{\sqrt{1-k\widetilde{r}^2}} = a\Big(r + \mathcal{O}(r^3)\Big)\,, \tag{1.34}$$

since the integral gives either r, $\arcsin(r)$ or $\operatorname{arcsinh}(r)$ depending on k, and all three of these equal r up to second order.

Do note that we neglected the temporal part of the metric: this is equivalent to assuming that photons travel instantaneously. So, this is Newtonian and rough, but it gives us the correct intuitive idea. We now wish to make this reasoning more precise.

The first step, since we want to discuss our observations of light, is to define the redshift and the luminosity distance.

Definition 1.3.1 (Redshift)**.** *The redshift z of a photon is defined by*

$$z = \frac{\lambda_0 - \lambda_e}{\lambda_e} = \frac{\lambda_0}{\lambda_e} - 1 = \frac{\nu_e}{\nu_0} - 1\,, \tag{1.35}$$

where λ_0 and λ_e are the observed and emission wavelengths respectively, while ν are frequencies with the same notation.

We will show that the redshift can be found from the ratio of the scale factors now and at emission: $1 + z = a_0/a_e$. Therefore, $\nu_o/\nu_e = a_e/a_0$.

We wish to study the distribution of light from an astronomical source: in Minkowski spacetime the apparent luminosity ℓ decreases like r^{-2} if r is the distance from the object. In a generic spacetime this will not be the case: however, we can define a measure of spatial distance d_L such that $\ell = L/(4\pi d_L^2)$, where L is the intrinsic luminosity of the object.

Do note that ℓ is dimensionally a luminosity flux: it is measured in units of energy per unit time per unit area.

Definition 1.3.2. *The luminosity distance d_L is defined as:*

$$d_L = \sqrt{\frac{L}{4\pi\ell}}\,. \tag{1.36}$$

Since $L \propto \ell$, this is a well-defined measure of distance between two generic points in spacetime, regardless of the presence of a source of light there.

How do we relate the luminosity distance and the scale factor? The radiation from our source is spread on a sphere: we integrate the angular part of the FLRW metric over a sphere of fixed comoving radius r to find its area.

The metric restricted to the angular coordinates at fixed r is given by

$$\mathrm{d}s^2 = a^2 r^2 \left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right), \tag{1.37}$$

so the area form on the surface of the sphere is

$$\mathrm{d}A = \sqrt{\det g}\,\mathrm{d}\theta \wedge \mathrm{d}\varphi\ , \tag{1.38}$$

where $\sqrt{\det g} = \sqrt{a^4 r^4 \sin^2\theta} = a^2 r^2 \sin\theta$. So,

$$A = \int_{S^2} \mathrm{d}A = a^2 r^2 \int_0^\pi \mathrm{d}\theta \int_0^{2\pi} \mathrm{d}\varphi \sin\theta = 4\pi a^2 r^2\,, \tag{1.39}$$

where we substituted the wedge product for the regular tensor product of the differentials since our axes are orthogonal. Now, at which time do we compute the scale factor? We are measuring the flux at the surface of the sphere, at the time at which we are observing: therefore we need to compute the scale factor at observation time as well. This gives us $A = 4\pi r^2 a_0^2$.

Now, the emitted luminosity is in the form:

$$L = \frac{\mathrm{d}N}{\mathrm{d}t_e} \langle h\nu_e \rangle\ , \tag{1.40}$$

where $\mathrm{d}N/\mathrm{d}t$ is the number of photons emitted per unit time, whose average energy is $\langle h\nu_e \rangle$. From the point of view of the observer, the number of photons is the same, while the frequency of the observed photon and the time interval $\mathrm{d}t_e$ change: specifically, $\nu_e = \nu_0 a_0 / a_e$ and $\mathrm{d}t_e = \mathrm{d}t_0 \, a_e / a_0$.[13] Therefore, the observed absolute luminosity obeys the relation $L = L_0 (a_0/a_e)^2$.

So, putting everything together we get:

$$\ell = \frac{L_0}{A} = \frac{L}{4\pi r^2 a_0^2} \left(\frac{a_e}{a_0}\right)^2 , \tag{1.41}$$

note that the value of k does not enter into the equation.

Therefore:

$$d_L = \sqrt{\frac{L}{4\pi \ell}} = \sqrt{\frac{4\pi r^2 a_0^2}{4\pi} \left(\frac{a_0}{a_e}\right)^2} = \frac{a_0^2}{a_e} r = a_0 (1+z) r \,. \tag{1.42}$$

Another way of defining a distance is the one we get by directly integrating the radial part of the line element: this way, we are effectively using a space-like measuring stick, working at fixed cosmic time and fixing both of the angles θ and φ.

Definition 1.3.3 (Proper distance)**.** *The proper distance d_P at a fixed cosmic time t to an object at a comoving radial coordinate r is given by:*

$$d_P = a(t) \int_0^r \frac{\mathrm{d}\widetilde{r}}{\sqrt{1 - k\widetilde{r}^2}} \,. \tag{1.43}$$

A derivation of the Hubble law

We want to derive the Hubble law ($v = H_0 d$) mathematically. It can also be restated as $cz = H_0 d$: the observed velocity of recession of the objects is measured through redshift, which is described (in the nonrelativistic limit[14]) by the formula

$$\lambda_0 = \lambda_e \left(1 \pm \frac{v}{c}\right) , \tag{1.45}$$

where the sign in the $\pm$ is a plus if the object is receding from us. We also defined z through $\lambda_0 / \lambda_e = 1 + z$: this means that we can identify $z = v/c$.

[13] This is the case even though the temporal component of the metric does not change, since we are not considering a fixed instance of cosmic time (which is unphysical): instead, we are considering the time intervals that the emitter and observed measure between the crests of a light wave sent from one to the other.

[14] The relativistic expression is

$$1 + z = \sqrt{\frac{1 + v/c}{1 - v/c}} \gtrsim 1 + \frac{v}{c} \,. \tag{1.44}$$

This will not be relevant for our discussion, however, since before the special-relativistic corrections become relevant we will need to use general-relativistic corrections: roughly speaking, the effects of spacetime expanding while the photon travels from its source to us.

Now, we are going to "move away from the current epoch by Taylor expanding": the scale factor at a time t, $a(t)$, can be written as

$$a(t) = a_0 + \dot{a}_0(t - t_0) + \frac{1}{2}\ddot{a}_0(t - t_0)^2 + \mathcal{O}\left(|t - t_0|^3\right) \tag{1.46a}$$

$$\approx a_0\left(1 + H_0(t - t_0) - \frac{1}{2}q_0 H_0^2 (t - t_0)^2\right), \tag{1.46b}$$

We drop the error term

Substituted $H_0 = \dot{a}_0 / a_0$.

where $q_0 \stackrel{\text{def}}{=} -\ddot{a}_0 a_0 / (\dot{a}_0)^2$ is called the *deceleration parameter* for historical reasons: people thought they would see deceleration ($\ddot{a} < 0$) when first writing this, so a positive q_0, but the deceleration parameter is instead measured to be negative.

Now, $1 + z = a_0 / a$ can be expressed as:

$$1 + z \simeq \left(1 + H_0(t - t_0) - \frac{1}{2}q_0 H_0^2 (t - t_0)^2\right)^{-1}. \tag{1.47}$$

Do note that this is derived starting with a formula which is correct up to second order in the time interval $\Delta t = t - t_0$: when expanding we cannot trust terms of order higher than second. Expanding with this in mind we get:[15]

$$1 + z \simeq 1 - H_0 \Delta t + \frac{q_0}{2} H_0^2 \Delta t^2 + H_0^2 \Delta t^2, \tag{1.52}$$

therefore

$$z = H_0(t_0 - t) + \left(1 + \frac{q_0}{2}\right) H_0^2 (t_0 - t)^2 \tag{1.53}$$

$$= (t_0 - t)\left[H_0 + \left(1 + \frac{q_0}{2}\right) H_0^2 (t_0 - t) z\right]. \tag{1.54}$$

Changed the time intervals to $t_0 - t = -\Delta t$, the square of which is the same as before.

Bringing the bracket to the other side we get

$$t_0 - t = z\left[H_0 + \left(1 + \frac{q_0}{2}\right) H_0^2 (t_0 - t)\right]^{-1} \tag{1.55}$$

[15] We need to compute the first and second derivatives of

$$\left(1 + H_0 \Delta t - \frac{q_0}{2} H_0^2 \Delta t\right)^{-1} = (1 + a\Delta t + b\Delta t)^{-1} = f(\Delta t) \ : \tag{1.48}$$

they are

$$\frac{\mathrm{d}f}{\mathrm{d}\Delta t} = -(1 + a\Delta t + b\Delta t)^{-2}(a + 2b\Delta t) \tag{1.49}$$

$$\frac{\mathrm{d}^2 f}{\mathrm{d}\Delta t^2} = 2(a + 2b\Delta t)^2 (1 + a\Delta t + b\Delta t) - (1 + a\Delta t + b\Delta t)^{-2}(2b), \tag{1.50}$$

so we have

$$\left.\frac{\mathrm{d}f}{\mathrm{d}\Delta t}\right|_{\Delta t = 0} = x \qquad \text{and} \qquad \frac{1}{2!}\left.\frac{\mathrm{d}^2 d}{\mathrm{d}\Delta t^2}\right|_{\Delta t = 0} = \frac{1}{2!}\left(2a^2 - 2b\right) = a^2 - b\,. \tag{1.51}$$

$$= z\left[H_0 + \left(1 + \frac{q_0}{2}\right)H_0 z\right]^{-1} \tag{1.56}$$

$$= \frac{z}{H_0}\left[1 + \left(1 + \frac{q_0}{2}\right)z\right]^{-1}, \tag{1.57}$$

where we substituted the first order expression $t_0 - t = z/H_0$: we are allowed to make this substitution since the expression is multiplied by z, which has the same asymptotic order as $t_0 - t$ (since H_0 is finite): working to first order inside the brackets is equivalent to working to second order in the global expression.

By the same reasoning, we can expand the inverse bracket to first order:

$$t_0 - t = \frac{z}{H_0}\left[1 - \left(1 + \frac{q_0}{2}z\right)\right] = \frac{z}{H_0} - \left(1 + \frac{q_0}{2}\right)\frac{z^2}{H_0}, \tag{1.58}$$

We would like the time interval to disappear: we want a distance, not a time, so we should seek an expression for r instead of Δt. We are observing photons, for which $\mathrm{d}s^2 = 0$, which is equivalent $c^2\,\mathrm{d}t^2 = a^2(t)\,\mathrm{d}r^2\,/(1 - kr^2)$. Taking a square root and integrating we get:

$$\int_t^{t_0} \frac{c\,\mathrm{d}t}{a(t)} = \pm \int_r^0 \frac{\mathrm{d}\widetilde{r}}{\sqrt{1 - k\widetilde{r}^2}}, \tag{1.59}$$

where we should select the negative sign since we want positive quantities on both sides. The other choice would correspond to the photon being emitted from the Earth and received at the comoving radius of the source.

The integral on the right hand side can be solved analytically: it is

$$-\int_r^0 \frac{\mathrm{d}\widetilde{r}}{\sqrt{1 - k\widetilde{r}^2}} = \begin{cases} \arcsin r = r + \mathcal{O}(r^3) & k = 1 \\ r & k = 0 \\ \operatorname{arcsinh} r = r + \mathcal{O}(r^3) & k = -1 \end{cases} \tag{1.60}$$

in all cases, it is just r up to *second* order (since the next term in the expansion of an arcsine or hyperbolic arcsine is of third order).

On the left hand side, we can substitute in $a(t)$ from equation (1.46b):

$$\int_t^{t_0} \frac{c\,\mathrm{d}t}{a(t)} = \frac{c}{a_0}\int_t^{t_0} \mathrm{d}\widetilde{t}\left[1 + H_0(\widetilde{t} - t_0) + \mathcal{O}(\Delta\widetilde{t}^2)\right]^{-1} \tag{1.61}$$

$$= \frac{c}{a_0}\left[t_0 - t + \frac{H_0}{2}(t - t_0)^2\right] + \mathcal{O}(\Delta t^3) \tag{1.62}$$

where we used the expression for the scale factor to first order only since the integration raised the order of the estimate by one. we have:

$$\frac{c}{a_0}\left((t_0 - t) + \frac{1}{2}H_0(t_0 - t)^2 + \mathcal{O}(\Delta t^3)\right) = r, \tag{1.63}$$

since the term proportional to q_0 only gives a third order contribution. We can now substitute the expression for the time difference with respect to the redshift (1.58), only computed to second order:

$$r = \frac{c}{a_0}\left[\frac{z}{H_0}\left(1 - \left(1 + \frac{q_0}{2}\right)z\right) + \frac{H_0}{2}\left(\frac{z}{H_0}\left(1 - \left(1 + \frac{q_0}{2}\right)z\right)\right)^2\right] \tag{1.64}$$

$$= \frac{c}{a_0}\left[\frac{z}{H_0} - \left(1 + \frac{q_0}{2}\right)\frac{z^2}{H_0} + \frac{H_0}{2}\frac{z^2}{H_0^2}\right] \tag{1.65}$$

Ignored the third and higher order terms in the square.

$$= \frac{c}{a_0 H_0}\left[z - \frac{1}{2}(1 + q_0)z\right]. \tag{1.66}$$

Now, we can insert this expression for r into the formula for the luminosity distance (1.42):

$$d_L = a_0^2 \frac{r}{a} = a_0(1+z)r \tag{1.67}$$

$$= a_0(1+z)\frac{c}{a_0 H_0}\left(z - \frac{1}{2}(1+q_0)z^2\right). \tag{1.68}$$

As we expect, the term a_0 disappears: it is a bookkeeping parameter, the physical properties of a universe described by a FLRW metric are invariant under a global rescaling of the scale factor.

Our expression also contains cubic terms in z: removing these to get back to second order we find

$$d_L = \frac{cz}{H_0}(1+z)\left(1 - \frac{1}{2}(1+q_0)z\right) \tag{1.69}$$

$$= \frac{cz}{H_0}\left(1 + \left(1 - \frac{1}{2} - \frac{1}{2}q_0\right)z\right) \tag{1.70}$$

$$= \frac{cz}{H_0}\left(1 + \frac{1}{2}(1 - q_0)z\right). \tag{1.71}$$

We can turn this into a relation for cz in terms of d_L by substituting in the first order expression $d_L H_0 = cz$ into the second order term:

$$cz = d_L H_0 + \frac{q_0 - 1}{2}cz^2 \tag{1.72}$$

$$cz = H_0\left(d_L + \frac{1}{2}(q_0 - 1)\frac{H_0}{c}d_L^2\right), \tag{1.73}$$

and we can notice that the relation is approximately linear and independent of acceleration for low redshift, but we can detect the acceleration at higher redshift. Typically we need to measure galaxies at least 10 Mpc away in order to detect these second order effects. As we mentioned in the beginning, the data show the parameter q_0 to be negative.

This effect is similar to a Doppler effect, but the analogy is not perfect: the redshift is caused by the expansion of space itself, and the apparent velocities of the galaxies at high redshift would be superluminal.

Interpretation of superluminal recession velocities

What follows is a synthesis of the enlightening article by Davis and Lineweaver [DL04], which should be referred to for clarification.

The proper way to define velocities is directly through the metric: so, we will need to differentiate the relation $d = a\chi$, where χ is the distance in comoving coordinates and d is the physical distance, with respect to the cosmic time. This yields $\dot{d} = \dot{a}\chi = Hd$. Note that we are assuming that the object is stationary with respect to the comoving coordinates ($\dot{\chi} = 0$), so we are ignoring what is called the *peculiar velocity*.

This $\dot{d}$ is precisely the recession velocity, and this definition can be extended to any redshift. As written it is quite implicit, since we do not know what the distance in comoving coordinates is to an object we observe with redshift z. This will be treated in section 2.4, but for now the result is (see also the aforementioned paper [DL04, eq. 1], and an older paper by Harrison [Har93, eq. 13]) that the velocity now of an object observed with redshift z is

$$v_{\text{rec}} = H_0 d_C(z) = cH_0 \int_0^z \frac{\mathrm{d}\tilde{z}}{H(\tilde{z})}, \tag{1.74}$$

which can be greater than c: doing the computation allows us to see that $v > c$ for $z \gtrsim 1.5$.[16]

Does this result contradict General, or even Special Relativity? It does not! In General Relativity we cannot directly compare velocities of objects which are far away from each other: they lie in different vector spaces, and there can be no local inertial frame extending that far. So, it is consistent to have superluminal recession velocities, while observers near the emission, as well as observers near Earth, always measure velocities locally to be $\leq c$.

Now we can clear some common misconceptions:

1. recession velocities can indeed exceed the speed of light;
2. they can do so in periods of "regular" expansion of the universe: we did not consider inflation, which surely did not occur for $z \sim 1.5$, which is the point at which the recession velocities start being superluminal;
3. we can indeed see the light from objects which are currently receding superluminally: the formula for the comoving distance is derived by integrating a photon's trajectory.

Definition 1.3.4 (Angular diameter distance)**.** *The angular diameter distance is defined as the ratio of the object's physical transverse size L to its angular size in radians $\Delta\theta$:*

$$d_A = \frac{L}{\Delta\theta} = a(t)r, \tag{1.75}$$

which is peculiar in that it is not monotonic in z [Hog00]: at $z \gtrsim 1$ it starts decreasing.

[16] An interesting (although meaningless) fact: the current recession velocity of the CMB ($z \approx 1090$) is around $3.14c \approx \pi c$, according to the latest Planck data [Col16]. For more details on how the recession velocity scales with the redshift, see figure 2 in Davis & Lineweaver [DL04].

Redshift-scale factor relation

Let us prove the statement from before, $\lambda_0/\lambda = a_0/a$: photons are emitted with a certain wavelength λ_e, at a comoving radius r from us, and detected at λ_o.

The line element for the photon is $\mathrm{d}s^2 = 0$, therefore $c\,\mathrm{d}t\,/a(t) = \pm\,\mathrm{d}r\,/\sqrt{1-kr^2}$.

As before, we can integrate this relation from the emission to the absorption: we call it $f(r)$ (it can be any of the functions shown in equation (1.60)):

$$\int_t^{t_0} = \frac{c\,\mathrm{d}\widetilde{t}}{a(\widetilde{t})} = f(r) \tag{1.76}$$

If we map $t \to t + \delta t$ and $t_0 \to \delta t_0$ in the integration limits, the integral must be constant since it only depends on r — do note that all the expansion of the universe is accounted for by the increasing scale factor, objects are stationary with respect to the comoving radial coordinate r. We are computing the integral for two successive wavefront of the light. Then, we equate the two:

$$f(r) = \int_t^{t_0} \frac{c\,\mathrm{d}\widetilde{t}}{a(\widetilde{t})} = \int_{t+\delta t}^{t_0+\delta t_0} \frac{c\,\mathrm{d}\widetilde{t}}{a(\widetilde{t})}\,, \tag{1.77}$$

which we can split into:

$$\left[\int_{t+\delta t}^{t} + \int_t^{t_0} + \int_{t_0}^{t_0+\delta t_0} - \int_t^{t_0}\right] \frac{c\,\mathrm{d}\widetilde{t}}{a(\widetilde{t})} = 0\,, \tag{1.78}$$

where, since all the integrals have the same argument, we collect it at the end for clarity. We simplify the original integral and swap the integration limits to get:

$$\int_t^{t+\delta t} \frac{c\,\mathrm{d}\widetilde{t}}{a(\widetilde{t})} = \int_{t_0}^{t_0+\delta t_0} \frac{c\,\mathrm{d}\widetilde{t}}{a(\widetilde{t})}\,, \tag{1.79}$$

which can we approximate by

$$\frac{c\delta t}{a(t)} = \frac{c\delta t_0}{a(t_0)}\,, \tag{1.80}$$

since the periods of the photons we are considering are generally much smaller than the cosmic timescales.

Since the frequency of the emitted and observed photons must be proportional to the inverse of the time intervals δt or δt_0, we have

$$\nu_e a(t_e) = \nu_o a(t_o)\,, \tag{1.81}$$

therefore

$$1 + z = \frac{\lambda_o}{\lambda_e} = \frac{a_0}{a}\,. \tag{1.82}$$

Chapter 2

Friedmann models

The Friedmann equations describe the dynamical evolution of the universe, as opposed to the static description given by the FLRW metric.

2.1 A Newtonian derivation of the Friedmann equations

The Friedmann equations are derived starting from the Einstein Field Equations for the FLRW metric, however we can derive them using an almost purely Newtonian argument.

We will not be able to recover the full equations, since a Newtonian fluid's pressure is $P \ll \rho c^2$: its contribution to the stress-energy-momentum tensor is negligible [TM20, eqs. 441–443]. So, through our argument we will recover the equations we would have with $P = 0$.

The only non-Newtonian step in our derivation is the justification of the Newtonian approximation: we wish to make use of the theorem attributed to Birkoff, but first derived by the Norwegian physicist J.T. Jebsen [JR05].

Proposition 2.1.1 (Jebsen-Birkhoff)**.** *The only solution to the vacuum Einstein Field Equations which is spherically symmetric is given by the Schwarzschild metric [MTW73, sec. 32.2]:*[1]

$$\mathrm{d}s^2 = \left(1 - \frac{2GM}{c^2 r}\right) c^2 \,\mathrm{d}t^2 - \left(1 - \frac{2GM}{c^2 r}\right)^{-1} - r^2 \,\mathrm{d}\Omega^2 \,. \tag{2.1}$$

With this in mind, let us take a spacetime with uniform density ρ. We consider a sphere, and imagine taking all the mass inside the sphere away.

By the Jebsen-Birkhoff theorem, the internal geometry of this shell is only determined by the mass distribution inside the shell: since we took all the mass inside the sphere away, the inside spacetime is Minkowski, that is, Schwarzschild spacetime with $M = 0$.[2]

[1] This is not relevant for what we will discuss here, but do note that there is no requirement of the generating (internal) mass distribution to be static: this is, in fact, the reason why spherically symmetric collapsing or pulsating stars cannot emit gravitational waves.

[2] There is actually some nuance to this: as is expounded upon in an article by Zhuang and Yi [ZY12], in general relativity the geometry inside the shell is actually influenced by the presence of the shell, in that its time coordinate is different to the one which would be measured by an outside observer. This is relevant, for

We describe this system through a mass coordinate: $M(\ell)$ is the mass enclosed by a spherical shell of radius ℓ.

The mass taken away will be $M(\ell) = \frac{4\pi}{3}\rho\ell^3$, where $\vec{l} = a(t)\vec{r}$ is the radius of the sphere, whose norm is $\ell = \left|\vec{l}\right|$, and ρ, as mentioned before, is the constant density.

We suppose that the gravitational field is *weak*: this is quantitatively expressed using the relation $\ell \ll r_g$, where r_g is the Schwarzschild radius of the system: this relation becomes

$$\frac{GM(\ell)}{\ell c^2} \ll 1\,, \tag{2.2}$$

and is a necessary assumption in order to apply a Newtonian approximation.

Now, we "put back" the mass which was removed, in order to restore the initial situation.

We put a test mass on the surface of the sphere. What is the motion of the mass due to the gravitational field from the center? It will surely be radial, and since as we said we are in the Newtonian approximation we can calculate it using Newton's equation:

$$\ddot{\ell} = -\frac{GM(\ell)}{\ell^2} = -\frac{4\pi G}{3}\rho\ell\,. \tag{2.3}$$

This seems to give us a net force even though we expect everything to be stationary because of isotropy. This is because we are not working in comoving coordinates: the radius of the sphere, ℓ, can change with the scale factor, even when the comoving vector $\vec{r}$ is constant.

Our final result will not depend on the unit vector we choose.

Plugging in $\ell = ar$ we find

$$\ddot{a}r = -\frac{4\pi G}{3}\rho a r \tag{2.4}$$

$$\ddot{a} = -\frac{4\pi G}{3}\rho a\,. \tag{2.5}$$

This is the second Friedmann equation, (1.9b), without the pressure term for the reasons mentioned before.

Starting from the acceleration equation (2.3) we can get

$$\dot{\ell}\ddot{\ell} = -\frac{GM}{\ell^2}\dot{\ell}\,, \tag{2.6}$$

Multiplied by $\dot{\ell}$

and identifying $2\dot{\ell}\ddot{\ell} = \mathrm{d}(\dot{\ell})^2 \big/ \mathrm{d}t$ we find the conservation of energy equation:

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left(\dot{\ell}\right)^2 = -\frac{GM\dot{\ell}}{\ell^2} \tag{2.7}$$

$$\frac{1}{2}\left(\dot{\ell}\right)^2 = -GM\int \frac{1}{\ell^2}\frac{\mathrm{d}\ell}{\mathrm{d}t}\,\mathrm{d}t \tag{2.8}$$

We integrate on both sides in $\mathrm{d}t$

instance, if we wish to compute the Shapiro delay of a ray of light passing through the shell and coming back out. This is not an issue for us, since the geometry on the inside is locally indistinguishable from a pure vacuum Minkowski spacetime if we measure only inside the shell.

$$\frac{1}{2}\dot{\ell}^2 = \frac{GM}{\ell} + C = \frac{4\pi}{3} G\rho\ell^2 + C\,, \tag{2.9}$$

Used $M = 4\pi\rho\ell^3/3$.

where C is an arbitrary constant, which we can express in terms of the scale factor:

$$\dot{a}^2 r^2 = \frac{8\pi G}{3}\rho a^2 r^2 + C \tag{2.10}$$

or, removing the r^2 term, which is a constant (since it is a comoving radius, with respect to which objects are stationary),

$$\dot{a}^2 = \frac{8\pi G}{3}\rho a^2 + \frac{C}{r^2}\,. \tag{2.11}$$

Now, the dimensions of this constant are those of a speed, therefore we can express it as $C/r^2 = -k_N = -kc^2$, where k is dimensionless — we are allowed to do this since k_N, the Newtonian curvature constant, has the dimensions of an energy per unit mass, or equivalently a velocity squared.

This clarifies the statement that the magnitude of k is arbitrary: we get it by dividing by the dimensionless comoving radius, whose magnitude is indeed arbitrary, therefore we can normalize it however we wish: so we choose $|k| = 1$ or 0.

The equation we found is of the form:

$$E_{\text{kin}} + E_{\text{grav}} = -kc^2\,, \tag{2.12}$$

where E_{kin} and E_{grav} are the energies per unit mass in the form of either kinetic or gravitational energy. So we can directly see that k, in a sense, describes the *intrinsic* energy of a free, stationary (with respect to the comoving coordinates) test mass.

If it is positive ($k < 0$) then the particles have an intrinsic positive energy, which causes expansion, while if it is negative ($k > 0$) it is as if they were intrinsically gravitationally bound, which causes contraction.

We can also recover the third Friedmann equation

$$\dot{\rho} = -3\frac{\dot{a}}{a}\left(\rho + \frac{P}{c^2}\right) \approx -3\frac{\dot{a}}{a}\rho\,, \tag{2.13}$$

where the approximation as before is the Newtonian one, $P \ll \rho c^2$. In this case, however, we will be able to also recover the nonrelativistic pressure term if we account for conservation of energy and not just mass.

When deriving these equations relativistically the third one comes from the conservation of the temporal component of the stress-energy tensor ($\nabla_\mu T^{\mu 0} = 0$), i.e. the "conservation of energy",[3] so to find it in our Newtonian calculation we will need to consider the nonrelativistic equivalent of that equation, which is the first law of thermodynamics.

[3] This is actually *not* a conservation law as stated, it is not covariant: in order for it to be we would need to project it along a temporal Killing vector, which does not exist in cosmology. However, if we did have a temporal Killing vector ξ_μ the projection $\xi_\nu \nabla_\mu T^{\mu\nu} = 0$ would indeed be the conservation of energy.

Indeed, it is more correct to say that this equation is not a conservation law at all, but is instead just an expression of the geometric Bianchi identities $\nabla_\mu G^{\mu\nu} = 0$, which are written in terms of the Einstein tensor $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}R g_{\mu\nu}$; if the Einstein equations hold then $G_{\mu\nu}$ is proportional to $T_{\mu\nu}$, therefore the two statements are equivalent.

We will consider *ideal fluids*. The first law, assuming adiabaticity — which must be present, since net heat transfer in the universe would violate isotropy — states:

$$dE + P\,dV = 0\,. \tag{2.14}$$

We can write the total energy as the product of the energy density times the volume: $E = \frac{4\pi}{3}\rho c^2 a^3$, since the volume is $V = \frac{4\pi}{3}a^3$.

So, the first law reads:

$$0 = \frac{4\pi}{3}\left[d\left(\rho c^2 a^3\right) + P\,d(a^3)\right] \tag{2.15}$$

$$= c^2\rho\,d(a^3) + c^2 a^3\,d\rho + P\,d(a^3) \tag{2.16}$$

Divided through by $4\pi/3$

$$= 3\left(\rho + \frac{P}{c^2}\right)\frac{da}{a} + d\rho\,, \tag{2.17}$$

Divided through by c^2a^3, collected terms.

which is the third Friedmann equation, (1.9c): the only manipulation left to do is to apply the differentials, which are covectors, to the temporal vector d/dt in order to turn them into time derivatives.

Why were we able to recover the relativistic term this time? The completely non-relativistic approach to this would be to write $M = \frac{4\pi}{3}\rho a^3$, and to write down the equation for the conservation of mass alone. Indeed, this would yield the Friedmann equation without the P term.

The three Friedman equations are not independent: for example, the second one (1.9b) can be derived from the first and third.

This means that we can derive the full relativistic equations in this Newtonian context, using the derivations we have shown for the first and third equation, and then combining these to find the second.

Let us do this derivation explicitly: we differentiate the first Friedmann equation

$$\dot{a}^2 = \frac{8\pi G}{3}\rho a^2 - kc^2 \tag{2.18}$$

with respect to time to find

$$2\dot{a}\ddot{a} = \frac{8\pi G}{3}\dot{\rho}a^2 + \frac{16\pi G}{3}\rho\dot{a}a \tag{2.19}$$

We then substitute in the expression we have for $\dot{\rho}$ from the third equation:

$$\dot{\rho} = -3\frac{\dot{a}}{a}\left(\rho + \frac{P}{c^2}\right), \tag{2.20}$$

which gives us

$$2\dot{a}\ddot{a} = \frac{8\pi G}{3}\left[-3\frac{\dot{a}}{a}\left(\rho + \frac{P}{c^2}\right)\right]a^2 + \frac{16\pi G}{3}\rho\dot{a}a \tag{2.21}$$

$$\ddot{a} = -4\pi G\left(\rho + \frac{P}{c^2}\right)a + \frac{8\pi G}{3}\rho a \tag{2.22}$$

Dividing through by $2\dot{a}$

$$\frac{\ddot{a}}{a} = -\frac{8\pi G}{3}\left[\frac{3}{2}\left(\rho + \frac{P}{c^2}\right) - \rho\right] \tag{2.23}$$

Dividing through by a

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + 3\frac{P}{c^2}\right), \tag{2.24}$$

which is precisely equation (1.9b).

2.2 The equation of state

So, the equation system is underdetermined: we do not in fact have three independent Friedmann equations, but just two. The variables we want to find, however, are three: $P(t)$, $\rho(t)$, $a(t)$.

So, we have to make an assumption: we will assume our fluid is a *barotropic* perfect fluid, that is, one for which the pressure only depends on the density: $P \overset{!}{=} P(\rho)$.

Very often this equation of state will be linear: $P = w\rho c^2$, with a dimensionless constant w.[4] We will assume this relation to be true.

2.2.1 Common equations of state

A thing we will compute for the different equations of state is the adiabatic[5] speed of sound:

$$c_s^2 = \frac{\partial P}{\partial \rho} = \frac{\mathrm{d}P}{\mathrm{d}\rho} = wc^2\,. \tag{2.25}$$

We also will be able to tell what the evolution of the energy density is for a varying scale factor: this can be derived from the third Friedmann equation (1.9c): in this case it reads

$$\frac{\dot{\rho}}{\rho} + 3\frac{P}{\rho c^2}\frac{\dot{a}}{a} + 3\frac{\dot{a}}{a} = 0 \tag{2.26}$$

$$\frac{\dot{\rho}}{\rho} + 3(1+w)\frac{\dot{a}}{a} = 0 \tag{2.27}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\log\left(\rho a^{3(1+w)}\right) = 0 \implies \rho a^{3(1+w)} = \text{const}\,, \tag{2.28}$$

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

1. $w = 0$ is equivalent to $P \equiv 0$: this is what we get in the nonrelativistic limit, for $P \ll \rho c^2$, since there is no pressure this can be interpreted as a *dust*. In this case $\rho \propto a^{-3}$; also, we have $c_s^2 \ll c^2$.

[4] This is a latin w, not a greek ω: students historically call it "omega" for some reason.

[5] The speed of sound is usually computed for adiabatic transformations, since the transmission of sound is usually close to an adiabatic process. In our case, adiabaticity is embedded in the hypotheses made in the derivation of the Friedmann equations. So, we can calculate the derivative without worrying about the adiabaticity condition being respected since for the solutions we will consider it always will be.

2. $w = 1/3$ is what we get if we seek the pressure of radiation.[6] In this case we have $c_s = c/\sqrt{3}$, while the energy density goes like $\rho \propto a^{-4}$, since we get a factor a^{-3} from the volume expansion and another a^{-1} from the decrease of the energy of each photon due to redshift. Alternatively, from what was derived before we can see that the exponent in the powerlaw must be $-3(1 + 1/3) = -4$.

   So, for a radiation-dominated universe the total energy $E \propto \rho a^3 \propto a^{-1}$ is not conserved.

3. $w = 1$ is called *stiff matter*: it has $P = \rho c^2$ and $c_s = c$. This is an incompressible fluid: it is so difficult to set this matter in motion that once one does it travels at the speed of light. Now, $\rho \propto P \propto a^{-6}$.

4. $w = -1$ means that $P = -\rho c^2$. We cannot compute a speed of sound (it would be imaginary). Now ρ and p are constants, since they are proportional to a^0. This is the case of dark energy: we will show in section 2.5 that the effect of inserting a cosmological constant Λ into the Einstein equations has precisely this effect.[7]

So, we replace the third Friedmann Equation with $w = \text{const}$ and

$$\rho(t) = \rho_* \big(a(t)/a_*\big)^{-3(1+w)}, \tag{2.29}$$

where a_* and ρ_* are the scale factor and density at some chosen time.

If we substitute this expression into the second FE we get that gravity is attractive ($\ddot{a} < 0$) if and only if $w > -1/3$.

Throughout this section we worked as if we had a single type of cosmic fluid in the universe: this is not really the case, we have many of them, and they will be interacting, but it is a good first approximation to consider them as separate.

In this plot, a can be interpreted as the time, since their relation is monotonic. We could insert the spatial curvature k in the plot: it decreases, but slower than matter, since it appears in the first Friedmann equation with an exponent a^{-2}. We can find an effective $\rho(a)$ law for the curvature by defining an effective ρ_k for curvature with $H^2 = \frac{8\pi G}{3}(\rho + \rho_k)$, which implies $\rho_k = -3kc^2/(8\pi G a^2)$.

We can also express this in units of the critical energy density $\rho_c = 3H^2/(8\pi G)$: we find

$$\frac{\rho_k}{\rho_c} = \Omega_k = -\frac{kc^2}{H^2 a^2}. \tag{2.30}$$

Now, the dark energy in the universe is the most important component. It it dominant over matter, radiation, and also dominant over spatial curvature.

[6] This expression can be derived in different ways, one of which is to start from the fact that the stress energy tensor must be traceless since it is of the form $T_{\mu\nu} \sim \sum_i \rho u_\mu^{(i)} u_\nu^{(i)}$, where $u_\mu^{(i)}$ are the four-velocities of photons: their norm is zero, so we must have $g^{\mu\nu} T_{\mu\nu} = 0$, but also for a perfect fluid $T = T_\mu^\mu = \rho - 3P/c^2$. Another, perhaps more illustrative derivation was given in the General Relativity course [TM20, pag. 86-87].

[7] This is shown by interpreting the additional term in the EFE as an addition to the stress-energy tensor and interpreting it as a perfect-fluid tensor [TM20, eqs. 434-438].

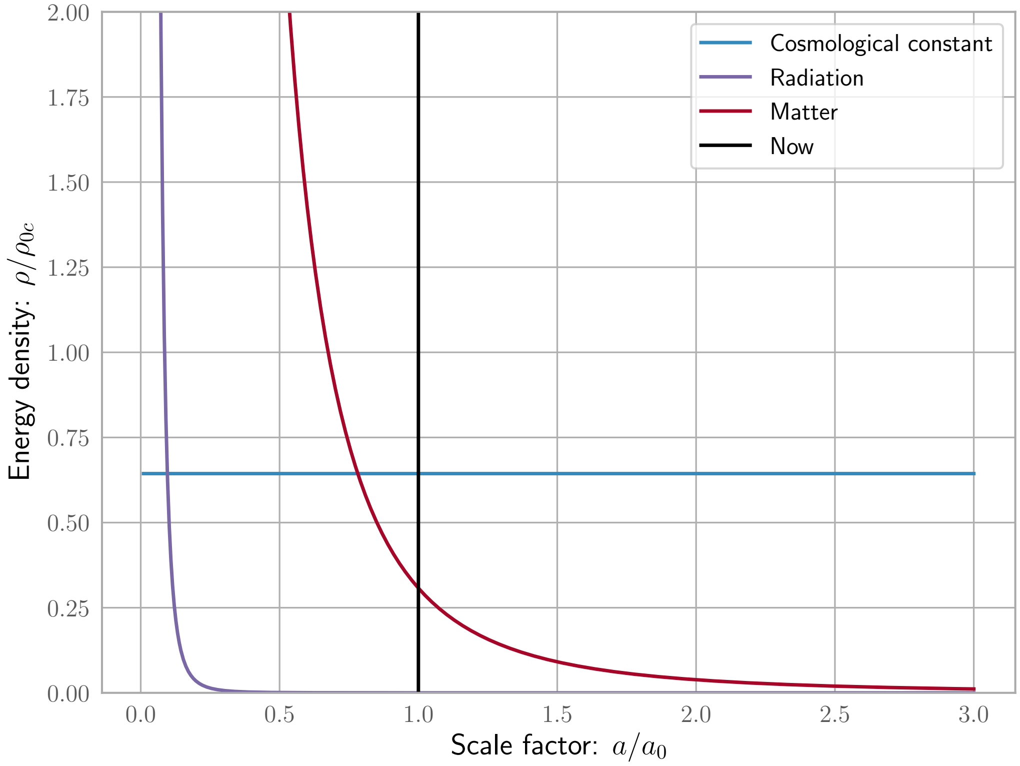

Figure 2.1: Contributions to the energy density varying with the scale factor. They are normalized to the current critical energy density, using data from the 2015 Planck mission [Col16]. The increase of the radiation energy density with an increasing scale factor is sharp, but the crossover point with the matter density is at $a = \rho_{\text{rad}}/\rho_{\text{m}} \approx 10^{-4}$, at which point we have $\rho \approx 10^{12}\rho_{0c}$.

2.3 Solutions of the Friedmann Equations

We want to solve the equation system

$$\frac{\dot{a}^2}{a^2} = \frac{8\pi G}{3}\rho - \frac{kc^2}{a} \tag{2.31}$$

$$\rho(t) = \rho_* \left(\frac{a(t)}{a_*}\right)^{-3(1+w)}, \tag{2.32}$$

which encompasses all of the physical content of the FE, since the second equation can be derived from these two.

Inserting (2.32) into (2.31) we find:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho_* \left(\frac{a}{a_*}\right)^{-3(1+w)} - \frac{kc^2}{a^2}. \tag{2.33}$$

2.3.1 Einstein-De Sitter models

As we discussed before, if we had zero spatial curvature ($k = 0$) then we would find $\rho = \rho_c = 3H^2/(8\pi G)$: so, we define the parameter $\Omega = \frac{8\pi G\rho}{3H^2} = \rho/\rho_C$ which quantifies how close this is to being true: experimentally this is compatible with 1.

The Einstein-de Sitter model is one where we take $\Omega \equiv 1$: negligible spatial curvature, which is equivalent to setting $k = 0$. So, the equation becomes:

$$\dot{a}^2 = \underbrace{\frac{8\pi G}{3}\rho_* a_*^{3(1+w)}}_{A^2} a^{-(1+3w)}\,, \tag{2.34}$$

Set $k = 0$, multiplied by a^2, defined A.

therefore $\dot{a} = \pm A a^{-\frac{1+3w}{2}}$, or $a^{\frac{1+3w}{2}}\,\mathrm{d}a = \pm A\,\mathrm{d}t$. We choose the positive sign, since we observe the universe to be expanding. This can be integrated directly: the equation is

$$\int_{a_*}^{a} \widetilde{a}^{\frac{1+3w}{2}}\,\mathrm{d}\widetilde{a} = A\int_{t_*}^{t}\mathrm{d}\widetilde{t} = A(t - t_*)\,, \tag{2.35}$$

but we must distinguish two cases: either $(1+3w)/2 = -1$, which is equivalent to $w = -1$, or not. Let us first assume that $w \neq -1$. Then, we get:

A solution is:

$$\frac{2}{3+3w}\, a^{\frac{3+3w}{2}}\bigg|_{a_*}^{a} = A(t - t_*) \tag{2.36}$$

$$a^{\frac{3+3w}{2}} - a_*^{\frac{3+3w}{2}} = \frac{3(1+w)}{2}\underbrace{\sqrt{\frac{8\pi G}{3}\rho_*}}_{H_*}\, a_*^{\frac{3+3w}{2}}(t - t_*) \tag{2.37}$$

Since $k = 0$ we have $H_*^2 = \frac{8\pi G}{3}\rho_*$

$$a^{\frac{3+3w}{2}} = a_*^{\frac{3+3w}{2}}\left(1 + \frac{3}{2}(1+w)H_*(t - t_*)\right) \tag{2.38}$$

$$a(t) = a_*\left(1 + \frac{3}{2}(1+w)H_*(t - t_*)\right)^{\frac{2}{3(1+w)}}\,, \tag{2.39}$$

which we can couple to the equation for the evolution of the density, by plugging this expression for the scale factor directly into (2.32):

$$\rho(t) = \rho_*\left(1 + \frac{3}{2}(1+w)H_*(t - t_*)\right)^{-2}\,, \tag{2.40}$$

and also the Hubble parameter $H \propto \sqrt{\rho}$:

$$H(t) = H_*\left(1 + \frac{3}{2}(1+w)H_*(t - t_*)\right)^{-1}\,. \tag{2.41}$$

There is a time where the bracket in $a(t)$ is zero, which means $a = 0$: this corresponds to the Big Bang, so we call it t_{BB}, defined by

$$1 + \frac{3}{2}(1+w)H_*(t_{\mathrm{BB}} - t_*) = 0\,. \tag{2.42}$$

Since the curvature scalar is $R \propto H^2 \propto \rho$,[8] at $t_{\rm BB}$ the curvature is diverges.
This time can be expressed by inverting the equation:

$$t_{\rm BB} = t_* - \frac{2}{3(1+w)H_*} \,. \tag{2.44}$$

Hakwing & Ellis proved that if $w > -1/3$ we unavoidably must have a Big Bang.
We can define a new time variable by

$$t_{\rm new} \equiv t - t_{\rm BB} = (t - t_*) + \frac{2}{3H_*(1+w)} \,. \tag{2.45}$$

Using this new variable the t_* simplifies, and we can just write:

$$a \propto t_{\rm new}{}^{\frac{2}{3(1+w)}} \,. \tag{2.46}$$

Inserting this new time variable (which we will just call t), we get that the Hubble parameter is:

$$H(t) = \frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}t} = \frac{2}{3(1+w)} t^{\frac{2}{3(1+w)}-1} t^{-\frac{2}{3(1+w)}} \tag{2.47}$$

$$H(t) = \frac{2}{3(1+w)t} \,, \tag{2.48}$$

so we can compute the density:

$$\rho(t) = \frac{3H^2}{8\pi G} = \frac{3}{8\pi G}\frac{4}{9(1+w)^2 t^2} \tag{2.49}$$

$$= \frac{1}{6(1+w)^2 \pi G t^2} \,. \tag{2.50}$$

Let us now revisit the cases from before:

1. $w = 0$ is nonrelativistic matter: it has $a \propto t^{2/3}$, $\rho = 1/(6\pi G t^2)$ and $H = 2/(3t)$.

   This yields a prediction for the age of the universe of $t \approx 9.6\,\mathrm{Gyr}$ (using the Planck data [Col16]): this is not correct, since the actual value is more like $t \approx 13.8\,\mathrm{Gyr}$, but it has the right order of magnitude; the discrepancy is due to the fact that the assumption of the universe being dominated by nonrelativistic matter is wrong.

2. $w = 1/3$ is radiation: it has $a \propto t^{1/2}$, $\rho \propto 3/(32\pi G t^2)$ and $H = 1/(2t)$;

[8] This can be shown by a simple argument: we take the trace of the EFE, to get

$$g^{\mu\nu}\left(R_{\mu\nu} - \frac{1}{2}R g_{\mu\nu}\right) = 8\pi G g^{\mu\nu} T_{\mu\nu} \implies -R = 8\pi G T \,, \tag{2.43}$$

where $R = g^{\mu\nu}R_{\mu\nu}$ and $T = g^{\mu\nu}T_{\mu\nu}$ are the traces of the Ricci tensor and of the stress-energy tensor. We've shown that $R \propto T$: so, since $T \propto \rho$, we have $R \propto \rho$, but if we also assume that there is no spatial curvature then $H^2 \propto \rho$, therefore $R \propto H^2$.

3. $w = -1$ is dark energy: as we mentioned before, this is the case which must be treated separately. The integral yields:

$$\log\left(\frac{a}{a_*}\right) = A(t - t_*)\,, \tag{2.51}$$

which means

$$a(t) = a_* \exp\left(H_* a_*^{\frac{3(1+w)}{2}}(t - t_*)\right), \tag{2.52}$$

while, as we saw, ρ and P (and, therefore, H) are constant.

From the redshift we can trace back the time of emission of the photon: for a matter-dominated universe, for example, we have:

$$1 + z = \frac{a_0}{a} = \left(\frac{t_0}{t}\right)^{2/3} = \left(\frac{2}{3H_0 t}\right)^{2/3}. \tag{2.53}$$

We can calculate the deceleration parameter for this case, to check that it is indeed positive: we need to differentiate the expression

$$a(t) = a_0 \left(\frac{3H_0 t}{2}\right)^{2/3}, \tag{2.54}$$

and we find

$$q_0 = -\frac{\ddot{a}_0 a_0}{\dot{a}_0^2} = -\frac{t_0^{-4/3} t_0^{2/3}}{(t_0^{-1/3})^2} \times \frac{\frac{2}{3}\left(-\frac{1}{3}\right)}{\frac{2}{3}\frac{2}{3}} = \frac{1}{2}\,. \tag{2.55}$$

The $a_0(3H_0/2)^{2/3}$ terms all simplify.

This is a special case of the fact that [LC02, eq. 2.2.4b]

$$q \equiv q_0 = \frac{1 + 3w}{2}\,. \tag{2.56}$$

2.4 Measuring distances

We want to be able to compute the comoving radius, given our knowledge of the evolution of the distribution of energy density in time.

We have shown that the luminosity distance is given by:

$$d_L \equiv \sqrt{\frac{L}{4\pi\ell}} = a_0(1 + z) r(z)\,. \tag{2.57}$$

Also recall *conformal time* η, which is defined by its relation to cosmic time, $a(\eta)\,\mathrm{d}\eta = \mathrm{d}t$: it allows us to write the FLRW metric as

$$\mathrm{d}s^2 = a^2(\eta)\left(c^2\,\mathrm{d}\eta^2 - \frac{\mathrm{d}r^2}{1 - kr^2} - r^2\,\mathrm{d}\Omega^2\right). \tag{2.58}$$

This is very important when we talk about zero-mass particles, with no intrinsic length scale: the photon, which is our primary tool for astrophysical observations, is one of these. This can be written in terms of the variable χ:

$$ds^2 = a^2(\eta)\left(c^2\,d\eta^2 - d\chi^2 - f_k^2(\chi)\,d\Omega^2\right), \tag{2.59}$$

where $f_k(\chi) = r$ is equal to $\sin(\chi)$, χ or $\sinh(\chi)$ if k is equal to 1, 0 or -1; in other words we either have $\chi = \arcsin(r)$, $\chi = r$ or $\chi = \text{arcsinh}(r)$.

If we look at photons moving radially we do not need to account for the angular part, and we find

$$ds^2 = 0 = a^2(\eta)\left(c^2\,d\eta^2 - d\chi^2\right), \tag{2.60}$$

therefore $c^2\,d\eta^2 = d\chi^2$: we get $c\big(\eta(t_0) - \eta(t_e)\big) = \chi(r_e) - \chi(r_0)$, where a subscript e means "emission", while a subscript 0 means detection. We are choosing the negative sign when simplifying the square, since the problem we are considering is that of radiation starting from an astrophysical source and coming towards us: its radial coordinate χ decreases when the temporal coordinate η increases.

This means that we can find out the comoving distance $\Delta\chi$ between two events by calculating the difference between their comoving times $\Delta\eta$. This is what was meant by the fact that this expression of the metric is useful for massless particles: the scale factor gets factored out, we can write the expression in a very simple way.

$$d\eta = \frac{dt}{a} = \frac{da}{a\dot{a}}, \tag{2.61}$$

and now recall $(1+z) = a_0/a$: we differentiate this with respect to time to find

$$\frac{dz}{dt} = -\frac{a_0}{a^2}\dot{a} = -\frac{a_0 H(z)}{a}, \tag{2.62}$$

which means

$$d\eta = \frac{dt}{a} = -\frac{dz}{a_0 H(z)}, \tag{2.63}$$

Took the inverse of the equation, split the differentials, used the definition of η

so we get our final expression:

$$d\chi = \frac{c\,dz}{a_0 H(z)}. \tag{2.64}$$

Used the fact that $d\chi = -c\,d\eta$.

So, if we can find a way to parametrize the Hubble parameter $H(z)$ in terms of the redshift we will be able to measure distances.

The Hubble parameter is given by

$$H^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2}, \tag{2.65}$$

where the density comes from several components: $\rho(t) = \rho_r(t) + \rho_m(t) + \rho_\Lambda$, where the first term is the density of radiation and scales like a^{-4}, the second is the density of matter and scales like a^{-3}, the third is the density of dark energy and is constant.

In terms of the redshift, they scale like $(1+z)^4$, $(1+z)^3$ (and $(1+z)^0$) respectively.

We express the Hubble parameter as a multiple of its value now: $H(z) = H_0 E(z)$, where $E(z)$ is a dimensionless function.

Recall the definition of $\Omega(t)$: it describes the ratio of the density of a certain type of fluid to the critical density. We can look at the $\Omega_i(t)$ for i corresponding to matter, radiation and so on:

$$\Omega_i(z) = \frac{8\pi G \rho_i(z)}{3H^2(z)} = \frac{8\pi G \rho_i(z=0)}{3H_0^2} \times \frac{\rho_i(z)/\rho_i(z=0)}{E^2(z)} = \Omega_{i,0} \frac{(1+z)^\alpha}{E^2(z)}, \tag{2.66}$$

where α is the exponent of the scaling of the fluid: $\alpha = 4$ for radiation, $\alpha = 3$ for matter, $\alpha = 0$ for the cosmological constant Λ, while for spatial curvature $\alpha = 2$.

For the Ω corresponding to the curvature we define: $\Omega_k = -kc^2/(a^2H^2)$ (see equation (2.30)).

We must have

$$1 = \Omega_r + \Omega_m + \Omega_\Lambda + \Omega_k. \tag{2.67}$$

We can write an expression for $E^2(z)$ by taking the ratio of the densities at emission versus now:

$$E^2(z) = \frac{H^2}{H_0^2} = \Omega_{\Lambda,0} + \Omega_{m,0}(1+z)^3 + \Omega_{r,0}(1+z)^4, \tag{2.68}$$

and to get E we just take the square root.

Now we can finally compute our integral

$$\chi(z) = \frac{c}{a_0 H_0} \int_0^z \frac{\mathrm{d}z'}{E(z')}, \tag{2.69}$$

therefore

$$r = f_k\left(\frac{c}{a_0 H_0} \int_0^z \frac{\mathrm{d}z'}{E(z')} \right). \tag{2.70}$$

This does depend on k, but the differences between positive and negative curvature are only relevant starting from third order. If the curvature is zero, we get the comoving distance:

$$d_C = r a_0 = \frac{c}{H_0} \int_0^z \frac{\mathrm{d}z'}{E(z')}. \tag{2.71}$$

If the curvature is not zero, we can still define a useful distance: the *transverse comoving distance*,

$$d_M = a_0 r = a_0 f_k\left(\frac{c}{a_0 H_0} \int_0^z \frac{\mathrm{d}z'}{E(z')} \right); \tag{2.72}$$

for $k = 0$ these two coincide.

Now, suppose we are looking at a certain far-away object with angular size $\Delta\theta$ and linear size *at emission* of Δx: then the *angular diameter distance* is given, in the small-angle approximation, by

$$d_A = \frac{\Delta x}{\Delta \theta} = a(t_e) r = \frac{a_0 r_z}{1+z} = \frac{d_M}{1+z}. \tag{2.73}$$

Distance name	Formula	Description
Comoving distance	$d_C = ra_0$ $= \frac{c}{H_0} \int_0^z \frac{\mathrm{d}z'}{E(z')}$	Distance in comoving coordinates multiplied by the current scale factor: if the expansion of the universe froze during our measurement, this is the distance we would measure between the two events. Assumes $k = 0$.
Transverse comoving distance	$d_M = ra_0$ $= a_0 f_k\left(\frac{c}{H_0 a_0} \int_0^z \frac{\mathrm{d}z'}{E(z')}\right)$	Generalization of the comoving distance to $k \neq 0$.
Luminosity distance	$d_L = d_M(1+z)$ $= \sqrt{L/(4\pi\ell)}$	Distance defined so that the radiative intensity we measure follows the inverse square law.
Angular diameter distance	$d_A = d_M(1+z)^{-1}$ $= \Delta x/\Delta\theta$	Distance defined by the ratio of a faraway object's size (measured using the scale factor at the time of the emission of the radiation we observe now) to its angular size.

Figure 2.2: A summary of the cosmological distances we defined, drawing on the summary by Hogg [Hog00].

Since the luminosity distance is given by

$$d_L = a_0(1+z)r = d_M(1+z) \tag{2.74}$$

their ratio is

$$\frac{d_L}{d_A} = (1+z)^2 . \tag{2.75}$$

2.5 The cosmological constant

Einstein thought that the universe had to be static: it was a common notion at the time that it should be, almost a philosophical principle.[9] Now we know that the universe is neither static nor stationary.[10]

So, he sought static solutions ($a = \text{const}$) for matter ($P = 0$) to the Friedmann equations (1.9): if we set $\dot{a} = \ddot{a} = 0$ the third equation becomes $\dot{\rho} = 0$, the second equation gives us

[9] An interesting historical fact: this was corroborated by a calculation error on Einstein's part, which was later pointed out by Friedmann. Einstein thought [Ein22] that $\nabla_\mu T^{\mu\nu} = 0$ implied $\partial_t \rho = 0$, while Friedmann pointed out [Fri22] that the correct equation reads $\partial_t\left(\sqrt{-g}\rho\right) = 0$: the density of the universe is not forced to be time independent if the determinant of the metric changes accordingly. Even the best make mistakes.

[10] The distinction between static and stationary is subtle but significant [Lud99]: *stationarity* is about the existence of a timelike Killing vector, while *staticity* is about the timelike Killing vector being orthogonal to spacelike submanifolds. A concrete example: Schwarzschild geometry is both static and stationary, Kerr geometry is stationary but not static, FLRW geometry is neither, since there is no timelike Killing vector field.

$\rho \equiv 0$, and from the first we must also have $k = 0$: the only way to have a static matter-filled universe is for the density of matter to be zero, and for the spatial curvature to be also zero.

In order to satisfy what he thought was an empirical fact, Einstein modified his equations in order to get a static non-empty solution.

The Einstein equations read

$$G_{\mu\nu} = 8\pi G T_{\mu\nu}\,, \tag{2.76}$$

when $c = 1$, where the Einstein tensor $G_{\mu\nu}$ can be defined in terms of the Ricci curvature tensor $R_{\mu\nu}$ and the scalar curvature R as:

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R\,. \tag{2.77}$$

This peculiar construction is the only one which can be made in terms of the curvature tensor and which is covariantly constant: $\nabla_\mu G^{\mu\nu} = 0$. This is a necessary condition since $\nabla_\mu T^{\mu\nu} = 0$: the Einstein equations state that they are proportional, so if we take the covariant derivative of the equations we must get the identity $0 = 0$.

Einstein added a term $-\Lambda g_{\mu\nu}$ to the LHS of the Einstein equations, with Λ a constant scalar. This is allowed since

1. it is tensorial (since it is a scalar multiple of the metric, which is a tensor);
2. it is symmetric;
3. it has zero covariant divergence, since Λ is constant and the metric is covariantly constant $\nabla_\mu g^{\mu\nu} = 0$.

Then, we can rewrite the EE in two equivalent ways: either

$$\widetilde{G}_{\mu\nu} = 8\pi G T_{\mu\nu} \qquad \text{with} \qquad \widetilde{G}_{\mu\nu} = G_{\mu\nu} - \Lambda g_{\mu\nu} \tag{2.78}$$

$$G_{\mu\nu} = 8\pi G \widetilde{T}_{\mu\nu} \qquad \text{with} \qquad \widetilde{T}_{\mu\nu} = T_{\mu\nu} + \frac{\Lambda g_{\mu\nu}}{8\pi G}\,. \tag{2.79}$$

In the first interpretation, the cosmological constant is an intrinsic geometric property of spacetime; in the second interpretation cosmological constant is a particular kind of fluid, with the property of its contribution to the stress-energy tensor always being a constant multiple of the metric.

In order to find out what the properties of this fluid are, w compare its stress-energy tensor to a generic ideal fluid tensor:

$$T^{(\text{generic})}_{\mu\nu} = \begin{bmatrix} \rho & 0 & 0 & 0 \\ 0 & P & 0 & 0 \\ 0 & 0 & P & 0 \\ 0 & 0 & 0 & P \end{bmatrix} \qquad T^{(\Lambda)}_{\mu\nu} = \frac{\Lambda g_{\mu\nu}}{8\pi G} = \frac{\Lambda}{8\pi G} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}\,, \tag{2.80}$$

so the corrections to the stress energy tensor must be $\rho \to \rho + \Lambda/8\pi G$ and $P \to P - \Lambda/8\pi G$, or, in other words, the density and pressure of the "cosmological constant fluid" are $\rho_\Lambda =$

$-P_\Lambda = \Lambda/8\pi G$. This proves that the equation of state of the cosmological constant is $w = -1$.

Inserting this into the Friedmann equations we get:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho + \frac{\Lambda}{3} - \frac{k}{a^2} \tag{2.81}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\rho + \Lambda \tag{2.82}$$

$$\dot{\rho} = -3\frac{\dot{a}}{a}\left(\widetilde{\rho} + \widetilde{P}\right) = -3\frac{\dot{a}}{a}\left(\rho + P\right), \tag{2.83}$$

and we can see that in the third equation, the effect of the source is encompassed in a term $\widetilde{\rho} + \widetilde{P}$: the two Λ terms cancel, since they are opposite. For a cosmological constant-dominated universe — that is, for a universe in which the only fluid behaves like the cosmological constant – we have $\dot{\rho} = \dot{P} = 0$.

So, proceeding with the derivation by Einstein, we set $\dot{a} = \ddot{a} = 0$: for the first Friedmann equation we get

$$\frac{8\pi G}{3}\rho + \frac{\Lambda}{3} = \frac{k}{a^2}, \tag{2.84}$$

and for the second:

$$4\pi G\rho = \Lambda. \tag{2.85}$$

So, we substitute the expression for $4\pi G\rho$ into the first Friedmann equation:

$$\frac{2}{3}(4\pi G\rho) + \frac{\Lambda}{3} = \Lambda\left(\frac{1}{3} + \frac{2}{3}\right) = \Lambda = \frac{k}{a^2}. \tag{2.86}$$

What are the physical conclusions to draw? Since we want matter in the universe we must have $\rho > 0$, which implies $\Lambda > 0$, which implies $k = 1$: so the universe must be closed.

Friedmann studied perturbations around this solution and found it to be unstable: so, it is not suitable as a description of the universe. This, combined with the observations by Hubble of an expanding universe, prompted the scientific community to discard the idea of a stationary universe in favor of an expanding one.

Einstein probably [Aut18] called the introduction of the cosmological constant into the equation his "greatest blunder"; however in modern cosmology the idea of a cosmological constant has gained new vigor: we observe the universe's expansion to be accelerated, that is $\ddot{a} > 0$, and the only way for this to be the case if $\rho > 0$ is if $\Lambda > 0$ as well. It is the only kind of fluid which has a repulsive gravitational effect.

As opposed to the approach by Einstein, in which the cosmological constant was inserted to stationarize the universe, we make it a measurable parameter of our theory.

A candidate for the cosmological constant term, which is a kind of intrinsic energy of space, is the vacuum energy in QFT: however the estimate we get when trying to make this quantitative is around 10^{120} times the measured value of Λ.

Friday 2010-10-18, compiled 2020-09-27

2.5.1 Evolution of a dark energy dominated universe

In order to find out how this parameter affects the universe's expansion, we consider a universe in which the only fluid behaves like the cosmological constant. So, we take the first Friedmann equation (2.81) in the absence of ordinary matter ($\rho = 0$) and with negligible spatial curvature ($k = 0$). This yields:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{\Lambda}{3}. \tag{2.87}$$

This is actually a good approximation for the asymptotic state of the universe, since the cosmological constant term is the only one which does not decay with the scale factor (and so, with time).

The solution to this differential equation is, as we mentioned in section 2.3:

$$a(t) = a_* \exp\left(\sqrt{\frac{\Lambda}{3}}(t - t_*)\right), \tag{2.88}$$

which can also be written as $a \propto e^{Ht}$, since $H = \dot{a}/a = \sqrt{\Lambda/3}$. This is called a steady-state solution, since the Hubble parameter is constant. It is also called a *de Sitter* solution: it belongs to the maximally symmetric 4D spacetime solutions to the Einstein Equations: Minkowski, de Sitter and Anti de Sitter: the latter has $\Lambda < 0$, the former has $\Lambda > 0$.

This actually seems to model the observed expansion of the universe well, and until recently it competed with the Big Bang theory.

The fraction of the cosmic fluid which behaves like dark energy is bound to increase with time, since as we saw it is the only component which does not decrease in density over time.

This is expressed formally using the so-called *no-hair cosmic theorem*, which is actually a conjecture if it is meant to describe the universe: it states that asymptotically only the dark energy contribution is relevant: all the matter and everything else is forgotten. In order to interpolate between the current — matter dominated, or in which at least matter has a sizeable contribution — universe and the asymptotic one we can use a solution in the form

$$a \propto \big(\sinh(At)\big)^{2/3}, \tag{2.89}$$

where we define $2A/3 = \sqrt{\Lambda/3}$, since the hyperbolic sine is asymptotically close to an exponential.

2.6 Curved models

We seek solutions to the Friedmann equations for nonzero spatial curvature k, for a universe containing nonrelativistic matter ($w = 0$) without dark energy. We make these assumptions since with them we can find an analytic solution.

We can rewrite the two independent Friedmann equations as

$$\dot{a}^2 = \frac{8\pi G}{3}\rho a^2 - k \tag{2.90}$$

$$\rho = \rho_0 \left(\frac{a}{a_0} \right)^{-3}, \tag{2.91}$$

and now we will solve them with $k = \pm 1$.

Solutions to parametric ODEs In general, for an ODE like $y = f(y')$ for the function $y = y(x)$ with f' continuous we introduce $y' \equiv p$, assuming $p \neq 0$: then $y = f(p)$, which implies

$$y' = \frac{\mathrm{d}f}{\mathrm{d}p} p', \tag{2.92}$$

Differentiated both sides of $y = f(p)$

which we can manipulate to get

$$p = \frac{\mathrm{d}f}{\mathrm{d}p} p' \implies \frac{\mathrm{d}x}{\mathrm{d}p} = \frac{1}{p} \frac{\mathrm{d}f}{\mathrm{d}p}, \tag{2.93}$$

so we can get the solution by integration: we get an expression for x in terms of p, which we will be able to invert since by assumption $p' \neq 0$: so, we get

$$x = \int \frac{1}{p} \frac{\mathrm{d}f}{\mathrm{d}p} \, \mathrm{d}p \qquad \text{and} \qquad y = f(p). \tag{2.94}$$

We use this for our problem: our differential equation looks like

$$\dot{a}^2 = \frac{8\pi G}{3} \rho a^2 - k \tag{2.95}$$

$$\dot{a}^2 = \frac{8\pi G}{3} \rho_0 \frac{a_0^3}{a^3} a^2 - k \tag{2.96}$$

$$\dot{a}^2 = A a^{-1} - k, \tag{2.97}$$

Substituted $\rho = \rho_0 a_0^3 / a^3$ from the third Friedmann equation.

where we defined $A \equiv 8\pi G a_0^3 \rho_0 / 3$. We can rewrite this as

$$a = \frac{A}{p^2 + k} = f(p) \qquad \text{where} \qquad p = \dot{a}. \tag{2.98}$$

Then, using the general formula we get:

$$t = \int \frac{1}{p} \frac{\mathrm{d}f}{\mathrm{d}p} \, \mathrm{d}p \qquad \text{where} \qquad \frac{\mathrm{d}f}{\mathrm{d}p} = - \frac{2Ap}{(p^2 + k)^2} \tag{2.99}$$

$$= \int \frac{-2A}{(p^2 + k)^2} \, \mathrm{d}p. \tag{2.100}$$

2.6.1 Positive curvature: a closed universe

If $k = +1$, then we can make the substitution $p = \tan(\theta)$, which is helpful since $1 + p^2 = \sec^2 \theta$; for the change of variable we have $\mathrm{d}p = \mathrm{d}\theta \sec^2 \theta$. So, for the time we find:

$$t = -2A \int \frac{\sec^2 \theta \, \mathrm{d}\theta}{\sec^4 \theta} \tag{2.101}$$

$$= -2A \int \cos^2(\theta)\,\mathrm{d}\theta = -A\big(\theta + \sin(\theta)\cos(\theta)\big) + \text{const}\,, \tag{2.102}$$

and we can apply the trigonometric identity $\sin(\theta)\cos(\theta) = \sin(2\theta)$:

$$t = -\frac{A}{2}\big(2\theta + \sin(2\theta)\big) + \text{const}\,. \tag{2.103}$$

Now we can define $2\theta = \pi - \alpha$, which allows for the simplification $\sin(2\theta) = \sin(\alpha)$; also, we can express $p = \tan\theta$ in terms of α. This gives us

$$t = \frac{A}{2}\big(\alpha - \sin(\alpha)\big) + \text{const} \qquad \text{and} \qquad p = \tan\left(\frac{\pi}{2} - \frac{\alpha}{2}\right). \tag{2.104}$$

Absorbed factor $-A\pi/2$ into the constant

We almost have our solution: inserting $p(\alpha)$ into the main equation for a (2.98) we get

$$a = \frac{A}{1 + \tan^2(\pi/2 - \alpha/2)} = A\cos^2\left(\frac{\pi}{2} - \frac{\alpha}{2}\right) \tag{2.105}$$

Used $1 + \tan^2 x = 1/\cos^2 x$.

$$= \frac{A}{2}\big(1 + \cos(\pi - \alpha)\big) = \frac{A}{2}\big(1 - \cos(\alpha)\big)\,, \tag{2.106}$$

Used $\cos^2(x/2) = (1 + \cos x)/2$ and $\cos x = -\cos(\pi - x)$.

which should be complemented with the equation we found for t: in the end, our solution looks like

$$t = \frac{A}{2}(\alpha - \sin\alpha) + \text{const} \tag{2.107}$$

$$a = \frac{A}{2}(1 - \cos\alpha)\,, \tag{2.108}$$

so, in order to interpret this physically we fix $t = 0 \iff \alpha = 0$, which sets "const" to zero, and we reinsert the constants. In order to do so, we wish to express the constant $A/2$ in term of observables such as H_0 and $\Omega_0 = \rho_0/\rho_{0c}$. We have:

$$\Omega_0 = \frac{8\pi G\rho_0}{3H_0^2} \qquad \Longrightarrow \qquad A = \frac{8\pi G a_0^3 \rho_0}{3} = \Omega_0 H_0^2 a_0^3\,, \tag{2.109}$$

which we can simplify by making use of the first Friedmann equation, which reads:

$$H_0^2 = \frac{8\pi G}{3}\rho_0 - \frac{k}{a_0^2} \qquad \Longrightarrow \qquad 1 = \Omega_0 - \frac{1}{a_0^2 H_0^2} \qquad \Longrightarrow \qquad a_0^2 H_0^2 = \frac{1}{1 - \Omega_0}\,, \tag{2.110}$$

We are treating the case $k = 1$

so we can write $A/2$ in two different ways:

$$\frac{A}{2} = \frac{a_0}{2}\frac{\Omega_0}{1 - \Omega_0} = \frac{1}{2H_0}\frac{\Omega_0}{(1 - \Omega_0)^{3/2}}\,. \tag{2.111}$$

We use one of these for a and the other for t: this is done because it makes the prefactor of the expression manifestly dimensionally consistent with the quantity we are expressing — this is not always the case when working with $c = 1$. Then, the expressions for a and t become:

$$a = a_0 \frac{\Omega_0}{2(\Omega_0 - 1)}\big(1 - \cos(\alpha)\big) = \widetilde{a}_0 \frac{1 - \cos\alpha}{2} \tag{2.112}$$

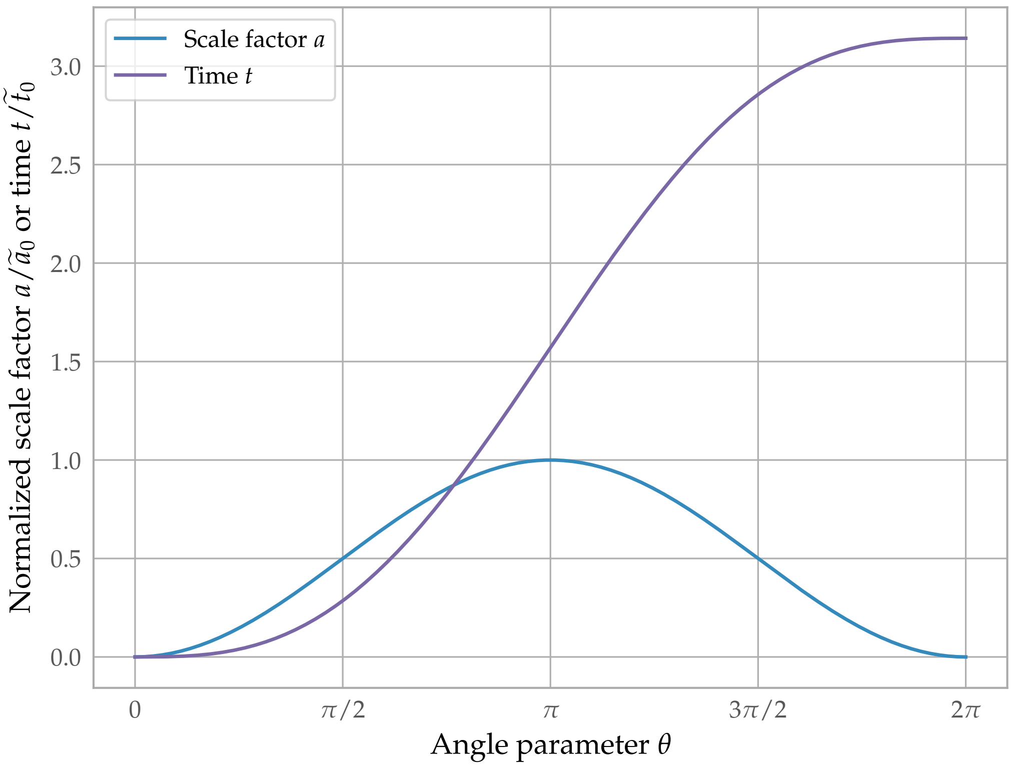

Figure 2.3: A plot of $a(\theta)$ and $t(\theta)$.

$$t = \frac{1}{H_0} \frac{\Omega_0}{2(\Omega_0 - 1)^{3/2}} (\alpha - \sin(\alpha)) = \widetilde{t}_0 \frac{\alpha - \sin \alpha}{2} \,. \tag{2.113}$$

For the discussion of these results we rename the angle variable from α to θ for historical reasons.

We have $\dot{a} > 0$ when $0 \leq \theta \leq \theta_m = \pi$, while $\dot{a} < 0$ when $\theta_m \leq \theta \leq 2\pi$: so, we call θ_m the *turn-around* angle. The angles 0 and 2π correspond to the Big Bang and the Big Crunch.

At θ_m we have:

$$a_m = \widetilde{a}_0 = a_0 \frac{\Omega_0}{\Omega_0 - 1} \tag{2.114}$$

$$t_m = \frac{\pi}{2} \widetilde{t}_0 = \frac{\pi}{2H_0} \frac{\Omega_0}{(\Omega_0 - 1)^{3/2}} \,. \tag{2.115}$$

The age of a closed universe

The total lifetime of the universe, in this scenario, is equal to $2t_m = \pi \widetilde{t}_0$. How does this compare to the result we found for a flat universe, namely $t_0 = 2/(3H_0)$ (equation (2.48) with $w = 0$)?

We set $a(t) = a_0$, which means we are normalizing the scale factor to the current one: this yields

$$1 = \frac{\Omega_0}{1-\Omega_0}\frac{1-\cos\theta}{2} \qquad \Longrightarrow \qquad \cos\theta = 1 - \frac{2(\Omega_0 - 1)}{\Omega_0} = \frac{2}{\Omega_0} - 1\,, \tag{2.116}$$

so we can invert the cosine (assuming we are in the expanding phase: it is not invertible globally) and insert our expression for θ into the expression for the time, to get[11]

$$t_0 = \frac{1}{2H_0}\frac{\Omega_0}{(\Omega_0-1)^{3/2}}\left(\arccos\left(\frac{2}{\Omega_0} - 1\right) - \frac{2}{\Omega_0}\sqrt{\Omega_0 - 1}\right). \tag{2.118}$$

As we can see in figure 2.4, this means that the estimated age of the universe is lower than $2/(3H_0) \approx 9.6\,\text{Gyr}$, while the measured age of the universe is around $14\,\text{Gyr}$.

2.6.2 Negative curvature: an open universe

For $k = -1$ we do exactly the same steps with hyperbolic functions instead of trigonometric ones, calling the argument of these functions ψ instead of θ: we get

$$a(\psi) = a\frac{\Omega_0}{2(1-\Omega_0)}\left(\cosh\psi - 1\right) \tag{2.119}$$

$$t(\psi) = \frac{1}{H_0}\frac{\Omega_0}{2(\Omega_0 - 1)^{3/2}}\left(\sinh\psi - \psi\right), \tag{2.120}$$

and as before we can calculate the independent variable with $\cosh\psi = 2/\Omega_0 - 1$.[12]

An analogous reasoning to the one before gives us

$$t_0 = \frac{1}{2H_0}\frac{\Omega_0}{(1-\Omega_0)^{3/2}}\left(\frac{2}{\Omega_0}\sqrt{1-\Omega_0} - \operatorname{arccosh}\left(\frac{2}{\Omega_0} - 1\right)\right), \tag{2.121}$$

Now $\sinh(\operatorname{arccossh}(x)) = \sqrt{x^2 - 1}$

which is plotted, again, in figure 2.4: in this case $t_0 > 2/(3H_0)$! This is then more attractive.

[11] We need to use the expression $\sin(\arccos(x)) = \sqrt{1-\cos^2(\arccos x)} = \sqrt{1-x^2}$ with $x = 2/\Omega_0 - 1$, and then the following manipulation:

$$\sqrt{1 - \left(\frac{2}{\Omega_0} - 1\right)^2} = \sqrt{1 - \frac{4}{\Omega_0^2} + \frac{4}{\Omega_0} - 1} = \frac{2}{\Omega_0}\sqrt{\Omega_0 - 1}\,. \tag{2.117}$$

[12] A doubt one might have: where does the sign change from $\theta - \sin(\theta)$ to $\sinh(\psi) - \psi$?

The difference between the calculations with the trigonometric functions and the hyperbolic functions lies in the substitution $2\theta = \pi - \alpha$: in the hyperbolic case we cannot do it this way, since the hyperbolic functions do not have any periodicity like this. Instead, the right substitution looks like $2\psi = i\pi + u$, since $\sinh(i\pi + u) = -\sinh(u)$.

Then, we have the same expressions as before, but their sign is flipped.

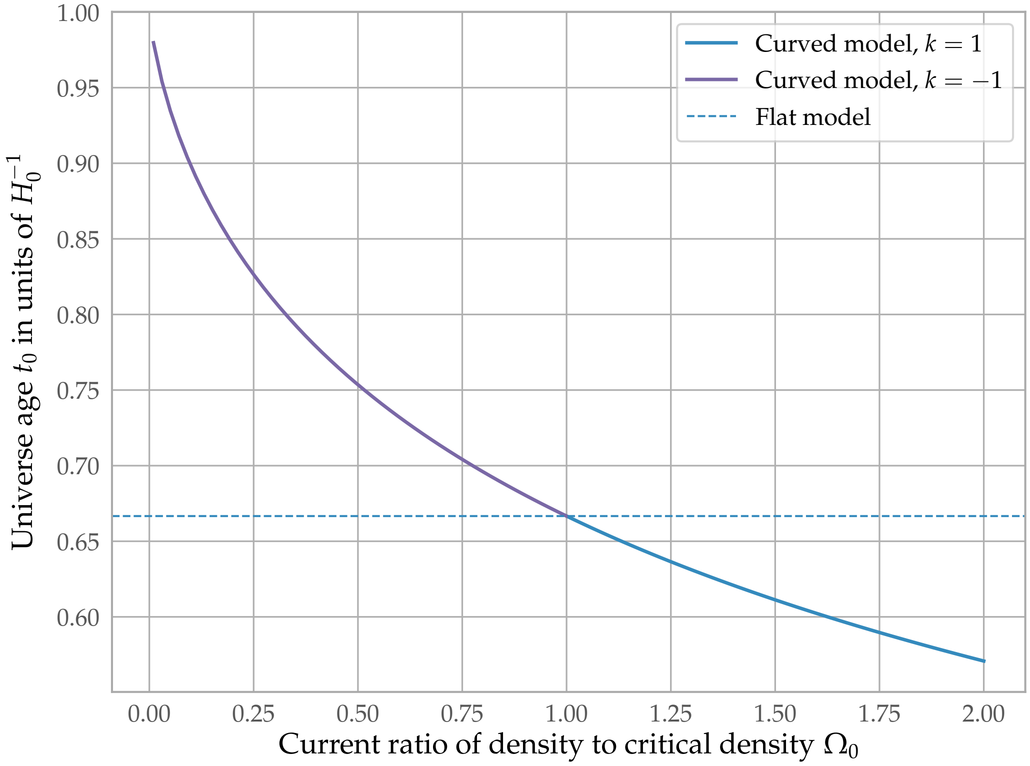

Figure 2.4: Universe age at the current time as a function of Ω_0. For $\Omega_0 > 1$ we use the positive curvature model in equation (2.118): the age is lower than $2/(3H_0)$; for $\Omega_0 < 1$ we use the negative curvature model (2.121): the age is greater than $2/(3H_0)$. The flat model is plotted with a horizontal line for clarity, but if the universe is spatially flat then we must have $\Omega_0 = 1$.

2.6.3 Considerations on curvature

The experimental fact that $t_0 > 2/(3H_0)$ seems to favour an open universe. However, the age of the universe is $t_0 \approx 0.96 H_0^{-1}$: looking at figure 2.4 it is clear that in order to account for it with spatial curvature only we would need $\Omega_0 \ll 1$, and actually $\Omega_0 < \Omega_{0m}$, where Ω_{0m} is the current measured ratio of the density of matter to the critical density.

In fact, in the current ΛCDM model of cosmology this is accounted for using dark energy, which means a positive cosmological constant.

From the second Friedmann equation

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\rho(1 + 3w) \tag{2.122}$$

we know that $\ddot{a} < 0$, that is, the expansion of the universe decelerates if $w > -1/3$. A singular instant at which $a = 0$ must be reached if $w > -1/3$, while this is not necessarily the case for $w < -1/3$. As we will discuss in the next chapter, this will be one of the motivations behind the theory of inflation, which does not require the presence of an initial singularity.

Chapter 3

The thermal history of the universe

In the study of the early stages of the universe the variation of the temperature, which determines the distribution of the energies of the collisions of the particles, plays a central role. It makes sense to talk about temperature when the particles are actually in thermal equilibrium, as they were in the early universe: photons and electrons were continuously Compton-scattering off each other. After the particles stop continuously interacting we say that they *decoupled*, and each component evolves independently.

The fact that the universe's temperature was much higher in the past is need to explain *primordial nucleosynthesis*: Helium-4 is the outcome of Hydrogen burning but in stellar evolution it is burned into heavier elements after it is formed, so we would expect to see small amounts of it. Instead, we see a relatively large amount of Helium-4: it makes up about a quarter of the universe by mass. The primordial universe being very hot helps account for this. In fact, this was first predicted in 1948, in the notorious $\alpha\beta\gamma$ paper [ABG48].

3.1 Radiation energy density and the equality redshift

In section 2.2.1 we discussed the evolution of the energy density of matter and radiation, showing that for radiation $\rho_r(z) = \rho_{0r}(1+z)^4$ while for matter $\rho_m(z) = \rho_{0m}(1+z)^3$. We discussed this in the context of electromagnetic radiation, but it describes well the behavior of very relativistic particles, such as neutrinos.

We can define a moment called the *equality redshift* z_{eq}. This is when the energy density of radiation and that of matter were equal: $\rho_r(z_{\text{eq}}) = \rho_m(z_{\text{eq}})$. This means that

$$\rho_{0,r}(1+z_{\text{eq}})^4 = \rho_{0,m}(1+z_{\text{eq}})^3 \implies (1+z_{\text{eq}}) = \frac{\rho_{0,m}}{\rho_{0,r}} = \frac{\Omega_{0,m}}{\Omega_{0,r}}\,, \tag{3.1}$$

where we divided and multiplied by the critical density today.

We know that $\Omega_{0,m}$ is around 0.3, while for the radiation we can deduce the density from the spectrum of the CMB.

Accounting for everything, we think that

$$1 + z_{\text{eq}} \simeq 2.3 \times 10^4 \Omega_{0,m} h^2 \approx 3370\,. \tag{3.2}$$

The value is that obtained from the Planck Collaboration [Col16].

This means that the recombination of electrons and protons into Hydrogen, which occurred around redshift $z_{\text{CMB}} \approx 1090$, happened when the universe was already *matter dominated* — specifically, the density of matter was ≈ 3 times that of radiation.

Another interesting time is z_Λ, when the energy density due to the cosmological constant equalled that of matter: $\rho_m(z_\Lambda) = \rho_\Lambda(z_\Lambda)$, which is calculated with the same reasoning as z_{eq}, recalling that ρ_Λ is a constant with respect to the redshift:

$$1 + z_\Lambda = \left(\frac{\rho_{0,\Lambda}}{\rho_{0,m}} \right)^{1/3} \simeq \left(\frac{0.7}{0.3} \right)^{1/3} \approx 0.33 \,. \tag{3.3}$$

This is relatively close, in cosmological terms: the comoving distance corresponding to this redshift is around 1350 Mpc, less than 10 % of the comoving distance to the CMB.

What is the temperature of a radiation-dominated universe? From the Stefan-Boltzmann law we know that $\rho_r \propto T^4$, while as we have discussed previously $\rho_r \propto a^{-4}$. Therefore, we expect $T \propto 1/a$ to hold: this is known as *Tolman's law*. In this chapter we will discuss how this does approximately hold, but we need to make some corrections due to the annihilation of ultrarelativistic particles.

We know that in a radiation dominated universe $a \propto t^{1/2}$, which means that $T \propto t^{-1/2}$.

We shall describe the pressure P, number density n and energy density ρ in the universe, as functions of the chemical potentials μ and of the temperature T. We will use natural units, so that $c = \hbar = k_B = 1$: so, temperatures and masses will be measured in electronVolts.

This is very convenient, since it allows us to make the following consideration: when the temperature will be of the order of the mass of a certain elementary particle, then statistically that type of particle will usually be ultra-relativistic.

This section will mostly follow Weinberg's book [Wei72, page 538, section 15.6].

3.2 Thermodynamics in the early universe

We will express the quantities mentioned above: P, ρ and n in terms of the distribution of particles in phase space: in general the phase space for a single particle in 3D is six-dimensional, but we operate under the assumption that the cosmological principle holds, so by homogeneity the spatial dependence of the distribution function can be neglected. Thus, we can talk of densities,[1] neglecting the spatial position, and integrating over momentum space to gather all the information there is to know about the particles in that position.

[1] Also, as a general rule, one should avoid talking about "global quantities": the universe is, in principle, infinite, so we should refer at most to what is or could be inside our light cone. It is better to work in terms of densities.

3.2.1 Number density, energy density and pressure

Number density

If $f(\vec{q})$ is the distribution density of particles with three-dimensional momentum $\vec{q}$, the number density of particles is given by:

$$n = \frac{g}{(2\pi)^3} \int \mathrm{d}^3 q \, f(\vec{q}, T, \mu) \,, \tag{3.4}$$

where the parameter g is the number of helicity states: it is the number of particles we can have with different quantum numbers, after fixing momentum and position.

This essentially is our choice for the normalization of the distribution function. We include the factor $(2\pi)^3$ in order to normalize the integral: the number of particles, a pure number, is given by

$$N \propto \int \mathrm{d}^3 q \, \mathrm{d}^3 x \, f(\vec{q}, \vec{x}) \,, \tag{3.5}$$

so the right hand side's differentials have the dimensions of an action cubed: we need to normalize them, and the conventional action used to do so is $h = 2\pi\hbar$. So, when we set $\hbar = 1$ we get a factor $(2\pi)^3$ on the denominator. Do note that the $\mathrm{d}^3 x$ integral, giving a volume, is brought to the left in our expression to give a number density.

The number of helicity states

The only quantum number which can vary after fixing those if we are considering an elementary particle is the spin component s_z; therefore if the total spin is s we should have $2s + 1$ possible spin states. For example electrons, which have spin $1/2$, will have $g = 2s + 1 = 2$.

Things, however, are more complicated than this: for photons, we only have two spin states ($g = 2$) even though they have $s = 1$, since $s_z = 0$ is unphysical for a photon. Gravitons also have $g = 2$ even though their total spin is $s = 2$: this is because $|s_z| \leq 1$ is unphysical for a graviton. In general for massless particles Lorentz invariance guarantees the fact that transverse modes cannot exist, since we cannot go in the rest frame of the particle.[2]

Even for massive particles we do not always have $g = 2s + 1$: g accounts for all internal degrees of freedom, and as the temperature drops below a certain value we need to consider composite particles as well: for atoms we also have vibration, rotation and such. These all contribute to g.

[2] Some more details on this: if we measure the spin component s_z to be equal to some value in a certain reference frame, then this will mean: (1) if $s_z \neq 0$ we need a rotation of at least $2\pi / s_z$ around the z axis in order to recover the system we started with, while (2) if $s_z = 0$ the system is symmetric with respect to rotations around the z axis.

So, for a photon to have $s_z = 0$ we would need to be in a frame in which its wavefunction was cylindrically symmetric. This cannot be the case if the photon is travelling in the z direction, so we must be in the rest frame of the photon, which does not exist.

Similarly, for gravitons the argument as to why $s_z \neq 0$ still applies, and we can exclude the spins $|s_z| = 1$ by the following argument: in full generality we can remove all the gauge freedom in a gravitational wave by going to TT gauge, and we can show that in TT gauge the wave is symmetric under rotations of angle π about the z axis. Therefore, the spin of the graviton must be at least 2 in magnitude.

Energy density

The energy density is given by

$$\rho = \frac{g}{(2\pi)^3} \int \mathrm{d}^3q \, E(q) f(\vec{q}, T, \mu) \,, \tag{3.6}$$

where $E^2 = q^2 + m^2$[3]. For photons $E = q$, for nonrelativistic particles $E \approx m + q^2/2m$. Here we are denoting the modulus of the momentum vector as $q = |\vec{q}|$.

This formula is a weighted average of the energies on the distribution function on the momenta.

Pressure

The adiabatic pressure is

$$P = \frac{g}{(2\pi)^3} \int \mathrm{d}^3q \, \frac{q^2 f(\vec{q}, T, \mu)}{3E(q)} \,. \tag{3.7}$$

This comes from a consideration of the diagonal components of the stress energy tensor of an ideal fluid: we know that

$$T^{\mu\nu} = \begin{bmatrix} \rho & 0 & 0 & 0 \\ 0 & P & 0 & 0 \\ 0 & 0 & P & 0 \\ 0 & 0 & 0 & P \end{bmatrix} = \frac{g}{(2\pi)^3} \left\langle \int \frac{\mathrm{d}^3q}{E(\vec{q})} f(\vec{q}) p^\mu(\vec{q}) p^\nu(\vec{q}) \right\rangle \,, \tag{3.8}$$

where $p^\mu = (E(q), \vec{q})$ is the momentum vector. Although it may not look like it, this formula is covariant.[a] The average sign is meant to indicate a spatial average, across volumes which are wide enough for homogeneity to hold.

This formula reproduces equation (3.6) for $\mu = \nu = 0$: indeed, in this case $p^0(q) = E(q)$. The off-diagonal components are zero by isotropy: if they were not, we would see heat and particle flow in specific directions.

So, for the diagonal components spatial components the integrand looks like $q^i q^j / E(q)$.

The formula for the pressure then follows by isotropy: the total force per unit area to go around is q^2/E, and it must be distributed equally in the three spatial directions, so if we want to switch from the directional integral $T^{ii} = P \propto \int q^i q^i / E$ (not summed over i) to an integral of the modulus of the momentum, $P \propto \int q^2/E$ we must divide by 3.

[a] we are integrating a tensorial expression ($f p^\mu p^\nu$ is a tensor) with respect to a covariant integration element: $\mathrm{d}^3q / E(q)$ is a scalar with respect to Lorentz transformations, since it can be obtained as

$$\mathrm{d}^4q \, \delta(E^2 - p^2 - m^2) = \frac{\mathrm{d}^3q}{2E} \delta(E - \sqrt{m^2 + p^2}) \,. \tag{3.9}$$

[3] Particles are on-shell, that is they obey the equations of motion (which is not mandatory, and this is what Quantum Mechanics is all about).

This definition gives us $P = \rho/3$ for photons directly, which can be seen by substituting $E = q$.

The distribution function

If the particles are in thermal equilibrium, the distribution in momentum space will be given by the following expression:

$$f(\vec{q}) = \left(\exp\left(\frac{E(q) - \mu}{T} \right) \pm 1 \right)^{-1}, \tag{3.10}$$

where we have a plus for fermions, and a minus for bosons. Here, $\mu = \partial\rho/\partial n$ is the chemical potential, the derivative of the energy (density) with respect to a change in the number (density) of particles: is becomes relevant when the gas becomes hot and dense, if it is sparse then adding particles does not affect the energy.

The Planck distribution, which describes the statistics of photons, is consistent with this, since it is given by:

$$f_k(\vec{q}) = \left(\exp\left(\frac{q}{T} \right) - 1 \right)^{-1}, \tag{3.11}$$

since they are bosons with no chemical potential.[4] The fact that the distribution of photons is indeed described by this distribution with $\mu = 0$ is a way to experimentally determine the fact that the chemical potential of photons is indeed zero. If we observed the distribution for physical blackbodies to have $\mu \neq 0$ this would be called a *spectral distortion*. The CMB is wonderfully consistent with $\mu = 0$, it is actually the best Planckian in Nature.

It is a fact that the chemical potential μ can be neglected when dealing with the early universe. Let us justify this.

[4] This may seem weird at first: is the Planck function not

$$B_\nu(T) = \frac{2\nu^3}{e^{2\pi\nu/T} - 1} \tag{3.12}$$

in natural units? Well, these two are actually equivalent formulations. To see this, recall that in natural units $q = E = \omega = 2\pi\nu$ for a photon. First of all, they describe different physical quantities: the Planck function describes the *spectral radiance*, $\mathrm{d}E = B_\nu(T)\,\mathrm{d}t\,\mathrm{d}A\,\mathrm{d}\nu\,\mathrm{d}\Omega$, while the distribution $f(q)$ describes the number density of particles per unit momentum volume: $\mathrm{d}N = f(q)\,\mathrm{d}^3q$.

To check their equivalence, let us compute the energy density with both:

$$\rho = \frac{g}{(2\pi)^3} \int \frac{q}{e^{q/T} - 1}\,\mathrm{d}\Omega\, q^2\,\mathrm{d}q = \frac{2}{(2\pi)^3} \int \frac{\omega^3}{e^{q/T} - 1}\,\mathrm{d}\Omega\,\mathrm{d}q = \int \frac{2\nu^3}{e^{2\pi\nu/T}}\,\mathrm{d}\Omega\,\mathrm{d}q \tag{3.13}$$

$$\text{but also} \quad \rho = \frac{\mathrm{d}E}{\mathrm{d}V} = \int \frac{\mathrm{d}E}{\mathrm{d}t\,\mathrm{d}A\,\mathrm{d}q\,\mathrm{d}\Omega}\,\mathrm{d}q\,\mathrm{d}\Omega = \int B_q(T)\,\mathrm{d}\Omega\,\mathrm{d}q\,, \tag{3.14}$$

where we used the fact that, in natural units $\mathrm{d}V = \mathrm{d}t\,\mathrm{d}A$.

In general, we can say that if for some chemical species we have the reaction $i + j \leftrightarrow k + l$, and we reach chemical equilibrium, then the chemical potentials of the species will be connected by $\mu_i + \mu_j = \mu_k + \mu_l$: this is called the **Saha equation**.

Assuming that we are in thermal equilibrium is not in general valid, we will do so in our discussion for simplicity, but non-equilibrium dynamics must be considered when dealing with CMB anisotropies. Although the assumption does not perfectly hold, this is quite instructive: the CMB spectrum is very close to an equilibrium blackbody spectrum, the deviations from equilibrium are small.

By enumerating all the possible chemical reactions between the various particle types we will get a system of equations for their chemical potentials, complemented with some known facts, such as the fact that photons have $\mu_\gamma = 0$, which is the case since they do not interact with each other.

For example, from the annihilation of electron and positron $e^+ + e^- \leftrightarrow 2\gamma$ we can derive a relation between the chemical potentials of e^+ and e^-: $\mu_{e^+} = -\mu_{e^-}$.

We can relate some chemical potentials by reactions, but not all of them: our system of equations will be degenerate, with degeneracy corresponding precisely to the globally conserved quantities (electric charge, lepton number, baryon number) which follow from the symmetry group of our theory. These can have any value and are conserved in any reaction,[5] so they cannot be fixed by the system.

If there was a global electric charge, we'd expect global magnetic fields, but we only see them with magnitudes of the order of the 1 nT, which gives an upper bound on the global charge of the universe. So, any global electric charge would be quite small — we will assume it is exactly zero.

We can estimate the orders of magnitude for the abundances of the various particle species in the universe. The baryon number is very small when compared to the number of photons in the universe, roughly speaking $n_\gamma / n_b \sim 10^{10}$.

The lepton number is harder to estimate, but it is reasonable to assume that it is quite small as well. For slightly more detailed discussion, see the book by Weinberg [Wei72, before eq. 15.6.5].

In the end, we can say that in the early universe $\mu / T \ll 1$, so we can assume $\mu \approx 0$. This is just a reasonable simplification, which we make in order to get analytic results.

Under this assumption the quantities characterizing the matter distribution in the universe only depend on the temperature: so, we will just write $n(T)$, $\rho(T)$ and $P(T)$.

In general, when dealing with thermodynamic problems in an expanding spacetime there is a complication: in Minkowski spacetime we have symmetry under time translations, thus it makes sense to talk about stationarity. In an expanding universe, instead, we have no Killing vector with respect to time. There is a competition between two evolutions, the thermodynamic evolution of the system and the expansion of the universe: we cannot truly have equilibrium!

[5] This holds as long as the temperature is low enough: we are considering the reactions which are allowed by the Standard Model of interactions, with its symmetry group $SU(3)_c \times SU(2)_L \times U(1)_Y$, but it is not currently known whether at higher temperatures (i. e. earlier times) this is the most general symmetry group which is spontaneously broken to the SM group. So, the statements we make only apply at relatively late times.

The way to deal with this problem is: we assume that the first evolution is much faster than the other, that is, we reach thermal equilibrium on timescales that are short if compared to the expansion. This way, we can neglect the expansion of the universe while our system reaches equilibrium.

So our problem is oversimplified: we assume thermodynamic equilibrium, which makes sense in certain periods of the life of the universe, and that allows us to embed a thermal situation into a universe which evolves in time.

3.2.2 Entropy

From the second principle of thermodynamics we know that the entropy in a certain volume V at temperature T, denoted $S(V, T)$ is given by:

$$\mathrm{d}S = \frac{1}{T}\left(\underbrace{\mathrm{d}\big(\rho(T)V\big)}_{\mathrm{d}E} + P(T)\,\mathrm{d}V\right) = \frac{1}{T}\big(V\,\mathrm{d}\rho(T) + (P(T)+\rho(T))\,\mathrm{d}V\big)\,, \tag{3.15}$$

since in order to get the total energy we must multiply the constant energy by the volume: $E = \rho(T)V$.

Then we can read off the partial derivatives of the entropy:

$$\frac{\partial S}{\partial V} = \frac{1}{T}\big(\rho(T)+P(T)\big) \qquad \text{and} \qquad \frac{\partial S}{\partial T} = \frac{V}{T}\frac{\mathrm{d}\rho(T)}{\mathrm{d}T}\,. \tag{3.16}$$

In order for the differential to be exact it needs to be closed, which means that the second partial derivatives need to commute (these are known as the *Pfaff relations*):[6]

$$\frac{\partial^2 S}{\partial T \partial V} = \frac{\partial^2 S}{\partial V \partial T} \tag{3.17}$$

$$\frac{\partial}{\partial T}\left(\frac{1}{T}\big(\rho(T)+P(T)\big)\right) = \frac{\partial}{\partial V}\left(\frac{V}{T}\frac{\mathrm{d}\rho(T)}{\mathrm{d}T}\right) \tag{3.18}$$

$$-\frac{1}{T^2}(\rho+P) + \frac{1}{T}\left(\frac{\mathrm{d}\rho}{\mathrm{d}T}+\frac{\mathrm{d}P}{\mathrm{d}T}\right) = \frac{1}{T}\frac{\mathrm{d}\rho}{\mathrm{d}T} \tag{3.19}$$

$$\frac{\mathrm{d}P}{\mathrm{d}T} = \frac{1}{T}(\rho+P)\,. \tag{3.20}$$

Simplified $T^{-1}(\,\mathrm{d}\rho/\mathrm{d}T\,)$, multiplied by T and brought $T^{-1}(\rho+P)$ to the other side

Cosmology has not entered into the picture yet, but it can by the third Friedmann equation, which can be rewritten as

$$\dot\rho = -3\frac{\dot a}{a}(\rho+P) \tag{3.21}$$

Multiplied by $-a^3$.

$$0 = 3\dot a a^2(\rho+P) + a^3\dot\rho \tag{3.22}$$

Added $a^3\dot P$ on both sides.

$$a^3\dot P = 3\dot a a^2(\rho+P) + a^3\dot\rho + a^3\dot P \tag{3.23}$$

$$a^3\dot P = \frac{\mathrm{d}(a^3)}{\mathrm{d}t} + a^3\frac{\mathrm{d}(\rho+P)}{\mathrm{d}t} \tag{3.24}$$

[6] They only hold in a simply connected space.

$$a^3 \dot{P} = \frac{\mathrm{d}}{\mathrm{d}t}\Big(a^3(\rho + P)\Big)\,, \tag{3.25}$$

and these two, when put together, are equivalent to

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{a^3}{T}(\rho(T) + P(T))\right) = 0\,, \tag{3.26}$$

therefore this quantity is a constant of motion. Let us verify this statement: expanding the derivative we get

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{a^3}{T}(\rho + P)\right) &= \frac{1}{T}\frac{\mathrm{d}}{\mathrm{d}t}\Big(a^3(\rho + P)\Big) + a^3(\rho + P)\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{T}\right) && (3.27)\\
&= \frac{1}{T}a^3\dot{P} - a^3(\rho + P)\frac{\dot{T}}{T^2} && (3.28)\\
&= \frac{a^3}{T}\frac{\mathrm{d}P}{\mathrm{d}T}\dot{T} - a^3(\rho + P)\frac{\dot{T}}{T} && (3.29)\\
&= \frac{a^3}{T}\frac{(\rho + P)}{T}\dot{T} - a^3(\rho + P)\frac{\dot{T}}{T} = 0\,. && (3.30)
\end{aligned}$$

Used equation (3.20).

For the RW line element, the square root of the determinant is given by $\sqrt{-g} = a^3$, so the conserved quantity can be written as

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\sqrt{-g}\frac{\rho + P}{T}\right) = 0\,. \tag{3.31}$$

This is relevant because the volume of any given spatial region scales with $\sqrt{-g}$ as the universe expands.

So the quantity which is differentiated is constant. If we plug this back into the differential expression for the entropy, we get:[7]

$$dS = d\left(\frac{(\rho+P)V}{T}\right), \tag{3.37}$$

therefore the differentiated quantities are equal up to an additive constant; from the conserved quantity we found and the fact that $V \propto a^3$ we now get that the **entropy is constant** in a comoving volume in thermal equilibrium:

$$S \equiv S(a^3, T) = \frac{a^3}{T}(\rho+P) = \text{const}\,. \tag{3.38}$$

Let us see what this entails: if we take photons, for example, we have $\rho \propto P \propto a^{-4}$: if we substitute this in we find that $a^{-4+3}/T = \frac{1}{aT}$ must be a constant, therefore $T \propto a^{-1}$. This is known as **Tolman's law**.

We only consider photons since they have a much larger number density.

3.2.3 Explicit expressions for the thermodynamic quantities

Let us give explicit expressions for the number density, energy density and pressure as a function of time. We are always assuming isotropy, so in all cases we will be able to simplify the angular part of the triple integral in d^3q as

$$\int d^3\vec{q} = 4\pi \int_0^\infty dq\, q^2\,, \tag{3.39}$$

so the three expressions will read

$$n(T) = \frac{g}{2\pi^2}\int dq\, q^2 f(q) \tag{3.40a}$$

[7] The procedure to prove this result is as follows: the expression we want to show is equal to dS can be written like

$$dS \stackrel{?}{=} d\left(\frac{(\rho+P)V}{T}\right) = \left(-\frac{(\rho+P)V}{T^2} + \frac{V}{T}\left(\frac{d\rho}{dT} + \frac{dP}{dT}\right)\right) dT + \frac{\rho+P}{T}\, dV\,, \tag{3.32}$$

while the definition of dS (3.15) can be written as

$$dS = \frac{V}{T}\frac{d\rho}{dT}\, dT + \frac{\rho+P}{T}\, dV\,, \tag{3.33}$$

so we can see that, since the term proportional to dV is the same in both cases, we only need to show that the coefficients of dT are equal, so what we need to prove is

$$-\frac{(\rho+P)V}{T^2} + \frac{V}{T}\left(\frac{d\rho}{dT} + \frac{dP}{dT}\right) \stackrel{?}{=} \frac{V}{T}\frac{d\rho}{dT} \tag{3.34}$$

$$-\frac{(\rho+P)V}{T^2} + \frac{V}{T}\frac{dP}{dT} \stackrel{?}{=} 0 \tag{3.35}$$

$$\frac{dP}{dT} = \frac{\rho+P}{T}\,, \tag{3.36}$$

which is precisely the statement we found to be equivalent to the Pfaff relations (3.20).

$$\rho(T) = \frac{g}{2\pi^2} \int \mathrm{d}q\, q^2 f(q) E(q) \tag{3.40b}$$

$$P(T) = \frac{g}{6\pi^2} \int \mathrm{d}q\, q^2 f(q) \frac{q^2}{E(q)}\,. \tag{3.40c}$$

In general these do not have analytic solutions, however if we only consider the ultrarelativistic and nonrelativistic limiting cases we can do the calculation.

Ultrarelativistic limit A particle being ultrarelativistic means that its momentum is much greater than its rest energy, $q \gg m$.

In our case we do not really care about any single particle being ultrarelativistic, rather, we ask that the temperature is high enough that the bulk of the particles is ultrarelativistic.

The momentum of any single particles will not always be large — in fact the distribution has its maximum at $q = 0$ — but the regions in which it is large give a much greater contribution than those in which it is small, as long as the temperature is large.

We define the rescaled momentum $x = q/T$, so that then the term appearing in the exponential is $E(q)/T = \sqrt{x^2 + m^2/T^2} \approx x = q/T$ under the assumption that $m/T \ll 1$.

With this assumption we get:

$$n(T) = \frac{g}{2\pi^2} \int_{\mathbb{R}^+} \mathrm{d}q\, q^2 \Big(\exp\big(q/T\big) \mp 1\Big)^{-1} \tag{3.41}$$

$$\rho(T) = \frac{g}{2\pi^2} \int_{\mathbb{R}^+} \mathrm{d}q\, q^3 \Big(\exp\big(q/T\big) \mp 1\Big)^{-1} \tag{3.42}$$

$$P(T) = \frac{g}{6\pi^2} \int_{\mathbb{R}^+} \mathrm{d}q\, q^3 \Big(\exp\big(q/T\big) \mp 1\Big)^{-1}\,, \tag{3.43}$$

so we can see that in this approximation, which is equivalent to $m \approx 0$, we get matter behaving like radiation: $P = \rho/3$.

The result of the integrals depends on the statistics of the particles (which determine the $\pm$ sign in the distribution), and it is given by the following expressions:

$$n(T) = \begin{cases} \dfrac{\zeta(3)}{\pi^2} g T^3 & \text{Bose-Einstein} \\[2ex] \dfrac{3}{4}\dfrac{\zeta(3)}{\pi^2} g T^3 & \text{Fermi-Dirac}\,, \end{cases} \tag{3.44}$$

where $\zeta(3)$ is the Riemann zeta function calculated at 3, giving

$$\zeta(3) = \sum_{n=1}^{\infty} \frac{1}{n^3} \approx 1.202\,. \tag{3.45}$$

and we get the proportionality to T^3. For the energy density:

$$\rho(T) = \begin{cases} \dfrac{\pi^2}{30} g T^4 & \text{Bose-Einstein} \\[2ex] \dfrac{7}{8}\dfrac{\pi^2}{30} g T^4 & \text{Fermi-Dirac}\,, \end{cases} \tag{3.46}$$

while to get the result for the pressure $P(T) = \rho(T)/3$ we just divide by 3.

We note that in natural units the Stefan-Boltzmann constant is $\sigma_{\text{SB}} = \pi^2/15$; the result we have found coincides with the Stefan-Boltzmann law for photons, which obey Bose-Einstein statistics and have two helicity states: $g = 2$, therefore

$$\rho(T) = 2\frac{\pi^2}{30}T^4 = \sigma_{\text{SB}}T^4 \,. \tag{3.47}$$

Nonrelativistic limit Now we work in the opposite limit, $m \gg T$. We can then expand the energy in powers of q/m (or T/m: as before, the point is that the typical value of q is T, so we can do it either way):

$$E = m\sqrt{1 + \frac{q^2}{m^2}} \approx m + \frac{q^2}{2m} + \mathcal{O}\left(\frac{q^2}{m^2}\right) . \tag{3.48}$$

The first temptation one might have is to work at the lowest possible order, approximating $E \approx m$. The exponential $\exp(E/T)$ will be very large compared to 1, so we can neglect the ± 1 in the denominator (which also means that the difference between bosons and fermions becomes negligible).

So, to zeroth order in (q/m) we get

$$f \approx \exp\left(-\frac{m-\mu}{T}\right) , \tag{3.49}$$

therefore the number density will be given by

$$n = \frac{g}{2\pi^2}\exp\left(-\frac{m-\mu}{T}\right)\int_{\mathbb{R}^+} \mathrm{d}q\, q^2 \,, \tag{3.50}$$

which diverges.

This is called the ultraviolet catastrophe: it is due to the fact that, while we are assuming q is small, we are not enforcing this in any way, and approximating all states as having the same energy regardless of their momentum. If, instead, we go to first order in q/m then we find

$$n = \frac{g}{2\pi^2}\exp\left(-\frac{m-\mu}{T}\right)\int_{\mathbb{R}^+} \mathrm{d}q\, q^2 \exp\left(-\frac{q^2}{2mT}\right) = g\left(\frac{mT}{2\pi}\right)^{3/2}\exp\left(\frac{\mu-m}{T}\right) , \tag{3.51}$$

where we applied the identity

$$\int_{\mathbb{R}} \mathrm{d}x\, x^2 \exp\left(-\alpha x^2\right) = \frac{\sqrt{\pi}}{2\alpha^{3/2}} \,. \tag{3.52}$$

The exponential factor $\exp(-m/T)$ is known as the Boltzmann suppression factor, which tells us that as long as relativistic and nonrelativistic particles are in thermal equilibrium there will be a much smaller number of the latter.

The energy density can be easily recovered from the number density if neglect higher order terms:

$$\rho(T) = \frac{g}{2\pi^2} \int \mathrm{d}q\, q^2 E(q) f(q) \approx \frac{g}{2\pi^2} \int \mathrm{d}q \left(m + \frac{q^2}{2m} \right) q^2 f(q) \approx m \underbrace{\frac{g}{2\pi^2} \int \mathrm{d}q\, q^2 f(q)}_{n(T)} . \tag{3.53}$$

For the pressure, on the other hand, we have

$$P(T) = \frac{g}{6\pi^2} \int \mathrm{d}q\, q^4 \frac{f(q)}{m + \frac{q^2}{2m}} \approx \frac{g}{6\pi^2} e^{-m/T} \int \mathrm{d}q\, \frac{q^4}{m} \exp\left(-\frac{q^2}{2mT} \right) , \tag{3.54}$$

and now we can apply the Gaussian integral identity [WA03, special case of eq. 10.1.11 (b)]:

$$\int_{\mathbb{R}} \mathrm{d}x\, x^4 \exp\left(-\alpha x^2 \right) = \frac{3}{8\alpha^2} \sqrt{\frac{\pi}{\alpha}} , \tag{3.55}$$

where for us $\alpha = 1/2mT$, which gives us

$$\begin{aligned} P(T) &= \frac{g}{6\pi^2} \frac{e^{-m/T}}{m} \frac{3(2mT)^2}{8} \sqrt{2mT\pi} = g \frac{m^{3/2} T^{5/2}}{\pi^{3/2} 2^{3/2}} e^{-m/T} = g \left(\frac{mT}{2\pi} \right)^{3/2} e^{-m/T} \times T & (3.56) \\ &= n(T) T . & (3.57) \end{aligned}$$

Therefore, $P = Tn = (T/m)\rho$, which tells us that the pressure of the nonrelativistic particles is much smaller than their energy density, since $T/m \ll 1$: we characterize them as *noninteracting dust*. The result we found, $P = nT$, is just the ideal gas law.

If we compare relativistic particles to nonrelativistic ones, the former dominate the latter in terms of all of these three quantities.

The physical context in which this becomes relevant, in the early universe, is that whenever the temperature drops below the mass of a certain particle, that particle starts to become nonrelativistic and its density drops exponentially, due to the Boltzmann suppression.

The main way for the particle to do so is generally to annihilate with its own antiparticle, thus producing radiation.

Effective degrees of freedom We have been discussing the behavior of a single particle species with g degrees of freedom; however we know that there were many types of particles in the early universe, so we need a way to generalize these results. We do so by defining the number of effective degrees of freedom:

$$g_*(T) = \sum_{i \in \mathrm{BE}} g_i \left(\frac{T_i}{T} \right)^4 + \frac{7}{8} \sum_{i \in \mathrm{FD}} g_i \left(\frac{T_i}{T} \right)^4 , \tag{3.58}$$

where the index i labels all the particle species in our model, running over all of those which are relativistic at temperature T (that is, as a first approximation, we only count those with masses $m_i < T$). The equilibrium temperature is T, while T_i are the temperatures of the various particle species, which we allow to be different from T — we will elaborate on this point in a moment. We distinguish two different terms in the sum, depending on whether the particle species obey Bose-Einstein or Fermi-Dirac statistics, since as we have seen the latter have a prefactor of 7/8 in the expression for the energy density.

This definition is constructed so that we can write the compact relation

$$\rho(T) = g_*(T)\frac{\pi^2}{30}T^4 \,. \tag{3.59}$$

Of course, considering particles completely when $m_i < T$ and not at all when $m_i > T$ is a simplification: in the region in which the temperature is of the order of the mass of the particle there will be a transition, which can be calculated properly by doing the integrals numerically. The results are shown in figure 3 of a paper by Husdal [Hus16], which can also be referred to for many more details on effective degrees of freedom. Figure 1 of the same paper shows how g_* decreases while the temperature of the universe decreases and more and more particle species become nonrelativistic.

Why do we consider the possibility of the temperature of a particle species being different from the equilibrium temperature?

Each process involving particles, be it decay or scattering, is characterized by a certain timescale. If the timescale of a certain interaction is larger than the cosmological timescale (the age of the universe), then that interaction statistically will not happen. Particles which cannot reach thermal equilibrium because of this are called *decoupled*, ones for which this is not the case are called *coupled*.

Although they may not interact, as long as they are relativistic decoupled particles can still affect the energy density of the universe, so we need to count them.

The time-temperature relation We want to find a relation between time and temperature in the early universe. Let us consider ultrarelativistic particles which are coupled, in the early universe which is radiation dominated (here "radiation" refers to all kinds of ultrarelativistic particles).

We start from the third Friedmann equation

$$H^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} \,, \tag{3.60}$$

neglect the curvature term[8] and use the facts that for a radiation-dominated universe $\rho \propto a^{-4}$ while $a \propto t^{1/2}$, meaning that $H = \dot{a}/a = 1/(2t)$.

Substituting these, as well as the expression we have found for the energy density in terms of the effective degrees of freedom, we get

$$\frac{1}{4t^2} = \frac{8\pi G}{3} g_*(T)\frac{\pi^2}{30}T^4 \,. \tag{3.61}$$

[8] We can do so since we know that right now the contribution to the global Ω of curvature is small (we have not been able to distinguish it from zero) and while this term scales as a^{-2} the matter term scales as a^{-3}. Since the matter term is dominant over the curvature term now, it was even more so earlier.

Then we have a formula for temperature in terms of time:

$$\frac{1}{2t} = \left(\frac{8\pi G}{3}\right)^{1/2} g_*^{1/2} \left(\frac{\pi^2}{30}\right)^{1/2} T^2 \tag{3.62}$$

$$t \approx \underbrace{\frac{1}{2\sqrt{\frac{8\pi\pi^2}{3\times 30}}}}_{\approx 0.301} g_*^{-1/2} \frac{m_P}{T^2} \approx \left(\frac{T}{\text{MeV}}\right)^{-2} \text{s}\,, \tag{3.63}$$

where $m_P = G^{-1/2} \approx 1.2 \times 10^{19}\,\text{GeV}$ is the Planck mass. Beware: there are different conventions for this, sometimes the definition is chosen as $m_P = (8\pi G)^{-1/2}$, which simplifies the Friedmann and Einstein equations somewhat. This mass corresponds to the energy scale at which quantum gravitational effects cannot be neglected.

The last approximation in (3.63) is quite rough, as it neglects the variation of the effective number of degrees of freedom completely: however, the factor $g_*^{-1/2}$ is of order 1 around $T \approx 1\,\text{MeV}$, which is the region in which we will apply our formula, so this is fine for our purposes.

Entropy effective degrees of freedom Entropy density is defined as entropy per unit volume, $s = S/V = (P+\rho)/T$. Since the total entropy in a comoving region is conserved (if there is thermal equilibrium) the quantity sa^3, proportional to S, is conserved.

If we only have relativistic particles (which satisfy $P = \rho/3$), the entropy density can be expressed as

$$s = (P+\rho)/T = \frac{4}{3}\frac{\rho}{T} = (2\pi^2/45) g_{*s} T^3\,; \tag{3.64}$$

where we defined a new number of effective degrees of freedom, g_{*s}, whose definition is slightly different from that of the one used for the energy, $s \propto T^3$ as opposed to $\rho \propto T^4$:

$$g_{*s} \equiv \sum_{i\in BE} g_i \left(\frac{T_i}{T}\right)^3 + \frac{7}{8}\sum_{i\in FD} g_i \left(\frac{T_i}{T}\right)^3 . \tag{3.65}$$

The expression for $s \propto g_{*s} T^3$ is more general than simply $s \propto T^3$, and in fact with this new one we can **update Tolman's law**: taking the cube root of the conserved quantity sa^3 we find $Tag_{*s}^{1/3} = \text{const}$.

3.2.4 Decoupling and radiation temperature

The temperature $1\,\text{MeV}$ occurs when the age of the universe is approximately $1\,\text{s}$, and this is the point at which the weak interactions involving neutrinos stop occurring.[9]

[9] This point is also relevant for another process: the weak interaction mediated processes are also what allows there to be an equilibrium between protons and neutrons, so when those reactions stop they become independent, and they evolve differently, since protons are stable while neutrons are not. This is crucial when discussing nucleosynthesis, the formation of nuclei.

When this happens, neutrinos decouple, so they stop interacting: they start evolving as any relativistic particle species would (they are relativistic since their mass is much lower than the MeV).

At this point, the temperatures of neutrinos and photons are "disconnected", there is no mechanism to equalize them. However, while the neutrinos are freely floating by, their energy density scaling like $\rho \propto a^{-4}$, the photons will be active for a few more seconds.

The mass of electrons and positrons is around 0.5 MeV, so until about 4 s into the life of the universe, 3 s from the decoupling of the neutrinos, the reaction $e^+ + e^- \leftrightarrow 2\gamma$ is still in equilibrium. After this, the populations of electrons and positrons annihilate, and since they have less effective degrees of freedom in which to deposit their energy than before (since the neutrinos are not coupled anymore) they dump it all into photons, thus increasing their temperature.

After this occurs, the photons keep evolving like any other relativistic particle species, with $\rho \propto a^{-4}$ but their temperature is higher than that of the neutrinos. Since they evolve in the same way, the ratio of the temperatures is constant. Now we will calculate this ratio.

We impose continuity of the entropy in a comoving volume across the transition which happens at 0.5 MeV between the stage in which electrons, positrons and photons are in equilibrium and the stage in which they decouple, since the electrons and positrons are not relativistic anymore and thus annihilate.

This transition only affects the temperature of the photons, while the neutrinos decoupled three seconds earlier; so, before the transition the temperatures of photons and neutrinos are equal, after it the photons' temperature increases.

Let us denote with an index $>$ the quantities pertaining to an earlier time, $t_{\text{transition}} > t$, while an index $<$ will denote the quantities pertaining to a later time, $t_{\text{transition}} < t$.

The "updated" version of Tolman's law reads $Tag_{*s}^{1/3} = \text{const}$, and if the transition is fast enough the scale factor can be taken to be equal on both sides of it. Therefore, we have

$$T_< = \left(\frac{g_{*s>}}{g_{*s<}} \right)^{1/3} T_> \,, \tag{3.66}$$

so we can compute the temperature after the transition, $T_<$, if we find the effective degrees of freedom before and after: before the transition we have photons, electrons and positrons. Photons have two polarization, and so do both electrons and positrons; also, the latter are fermions, so we find

$$g_{*s>} = 2 + \frac{7}{8}(4) = \frac{11}{2} \,, \tag{3.67}$$

while after the transition only photons are relativistic, so we have

$$g_{*s<} = 2 \,. \tag{3.68}$$

This means that the temperature of the photons increases by a factor

$$T_< = \left(\frac{11}{4} \right)^{1/3} T_> \approx 1.4 T_> \,, \tag{3.69}$$

which allows us to compute the neutrino temperature at any time, since $T_{<}/T_{>} = T_\gamma/T_\nu$, and because they scale in the same way:

$$T_\nu = T_\gamma \left(\frac{4}{11}\right)^{1/3} . \tag{3.70}$$

Right now, the temperature will be around $T_{0\nu} \approx (4/11)^{1/3} T_{0\gamma} \approx 1.94\,\mathrm{K}$, where $T_{0\gamma}$ is the current CMB temperature.

As an exercise, let us compute the number of effective degrees of freedom some time after the decoupling of electrons, say at $T = 0.1\,\mathrm{MeV}$. The global temperature T we are referring to is the one of the photons: so, applying the definition we find

$$g_* = \sum_{i\in BE} g_i \left(\frac{T_i}{T}\right)^4 + \frac{7}{8}\sum_{i\in FD} g_i \left(\frac{T_i}{T}\right)^4 \tag{3.71a}$$

$$= 2 + \frac{7}{8}\left(3\times 2\left(\frac{T_\nu}{T_\gamma}\right)^4\right) \tag{3.71b}$$

$$= 2 + \frac{21}{4}\left(\frac{4}{11}\right)^{4/3} \approx 3.36\,, \tag{3.71c}$$

since we need to consider neutrinos (of which there are three flavors, each having two polarization states), which contribute to the total energy density, but not electrons which are not relativistic anymore.

With this result, we can find the energy density of radiation at that time according to (3.59): we get

$$\rho_r(0.1\,\mathrm{MeV}) \approx 1.1\times 10^{-4}\,\mathrm{MeV}^4 = 25.7\,\mathrm{g cm}^{-3}\,. \tag{3.72}$$

Multiplied by $\hbar^{-3}c^{-5}$ to get the CGS units.

3.3 Problems with the Hot Big Bang model, inflation

Around the 1960s, cosmologies were trying to piece together a description of the early universe in terms of particle physics, as we discussed in this chapter up to here. However, soon it became apparent that the standard cosmological model in use has some inconsistencies. Let us now explore these.

3.3.1 The cosmological horizon problem

This was noticed as early as 1956. Let us consider radial null geodesics in a universe described by a FLRW metric. These are the worldlines of photons we can detect with telescopes. Imposing $\mathrm{d}s^2 = 0$ we find:

$$c^2\,\mathrm{d}t^2 = a^2(t)\frac{1}{1-kr^2}\,\mathrm{d}r^2\ , \tag{3.73}$$

which we can integrate (taking one of the two solutions for simplicity, choosing one over the other just amounts to parametrizing time in the opposite direction) to find

$$\int_0^t \frac{c\,\mathrm{d}t}{a(t)} = \int_0^r \frac{\mathrm{d}\widetilde{r}}{\sqrt{1-k\widetilde{r}^2}} = f(r)\,. \tag{3.74}$$

The function $f(r)$ gives us the proper distance between emission and detection of a photon, but it does so in terms of the adimensional coordinate r: in order to get something which has the dimensions of a length we need to multiply by a calculated at a certain time,[10]

$$d_{\mathrm{hor}}(t) = a(t)\int_0^t \frac{c\,\mathrm{d}\widetilde{t}}{a(\widetilde{t})}\,. \tag{3.75}$$

If this integral is convergent, we should be worried: let us see why.

If we integrate from the beginning of time to now, we get the spatial (current) comoving distance elapsed by a photon which started moving at the start of time. This is the radius of the largest region we could in principle observe. It is of the order of 3 Gpc. Since the integral is convergent this is finite, and it is increasing as time passes. So, ever-further regions are "coming into view" (at least in principle). Roughly speaking, the issue is that the regions which we start seeing at the edges should be causally disconnected from the ones already in view, so we would not expect them to exhibit the same properties — but *they do*. This is the basic idea, let us formalize it slightly and connect it to observations.

We cannot actually see light coming from the very edge of the in-principle-observable universe, since for redshifts larger than $z_{LS} \approx 1100$ the universe was opaque to electromagnetic radiation. The surface of points at this redshift is called the *Last Scattering* surface. So, we refer our expectations to the CMB, which was emitted as the primordial plasma became transparent.

The CMB was emitted at a cosmic time of $t \approx 3.8 \times 10^5$ yr after the Big Bang. It is observed to be very close to being uniform, with $\Delta T/T \sim 10^{-5}$ after correcting for the Doppler dipole modulation: it looks like a distribution emitted by matter in thermal equilibrium. Crucially, this holds at any angular scale we choose: the equilibrium is there across the whole sphere.

Recall that the *angular diameter distance* d_A is defined so that if an object with linear size at emission Δx spans an angle $\Delta\theta$ then we have (in the small-angle approximation)

$$d_A = \frac{\Delta x}{\Delta\theta}\,. \tag{3.76}$$

The angular diameter distance to the last scattering surface is approximately $d_A(z_{LS}) \approx 12.8$ Mpc. On the other hand, the scale of the particle horizon at that redshift can be calculated by taking the difference of comoving distances to us, $d_C(z_\infty) - d_C(z_{LS}) \approx 281$ Mpc and multiplying it by the scale factor, $a(z_{LS}) \approx 9.2 \times 10^{-4}$, which yields a horizon scale of

[10] Note that this choice is arbitrary: we are computing the comoving distance *as measured at the cosmic time of detection*.

approximately $r_H \approx 260\,\mathrm{kpc}$ at that time.[11]

We can then say that $\Delta x \sim r_H$, therefore the angular scale at which we expect to be able to observe correlations since there can be causal connections is around

$$\Delta\theta \approx \frac{\Delta x}{d_A} \approx 0.02\,\mathrm{rad} \approx 1.2^\circ\,. \tag{3.77}$$

A similar calculation [Toj, eqs. 8–12] yields $\Delta\theta \sim (1 + z_{LS})^{-1/2} \approx 1.7^\circ$, using the (reasonable) assumption of matter dominance in the epoch of recombination. Some steps there are not really clear to me, so I'm not sure whether my line of reasoning is equivalent (and valid) besides the assumption.

This is in stark opposition with the scale of observed correlations, which span the whole sky!

Mention in the lecture of the Mixmaster Universe by Misner (and the Bianchi classification of Lie Algebras for context) as an alternative to inflation — is this relevant here?

Cosmic inflation Now, if the quantity $d_{\mathrm{Hor}}(t)$ were to diverge this would mean that we could have a causal connection with any point in the universe, provided we went far enough back in time.

We can approximate

$$d_{\mathrm{Hor}}(t) = a(t) \int_0^t \frac{c\,\mathrm{d}\widetilde{t}}{a(\widetilde{t})} \approx ct \sim \frac{c}{H} \equiv d_{\mathrm{H}}\,, \tag{3.78}$$

where we defined the new *Hubble distance*, $d_H = c/H$. This is a physical distance, but we can also define the corresponding dimensionless comoving Hubble radius: $r_H = c/(Ha) = c/\dot{a}$, which satisfies $d_H = a r_H$.

We can hypothesize that there was a period in the early universe when the comoving radius r_H was decreasing with time: if this is the case, the regions we are observing today as "coming into view" could actually have been in causal contact in the early universe.

For the comoving radius to be decreasing ($\dot{r}_H < 0$), the condition is (neglecting factors of c, or working in natural units):

$$\dot{r}_H = -\frac{\ddot{a}}{\dot{a}^2} < 0\,, \tag{3.79}$$

[11] All the calculations were made automatically using the `astropy` package, using a flat ΛCDM model with parameters obtained from the Planck mission [Col16].

```python
from astropy.cosmology import Planck15 as cosmo
import numpy as np
import astropy.units as u
z_LS = 1089
dx = (cosmo.comoving_distance(np.inf) - cosmo.comoving_distance(z_LS)) * cosmo.scale_factor(z_LS)
dA = cosmo.angular_diameter_distance(z_LS)
(dx / dA).to(u.degree, equivalencies=u.dimensionless_angles())
```

therefore we need $\ddot{a} > 0$ for at least some time.

The second Friedmann equation (in natural units) tells us that

$$\ddot{a} = -\frac{4\pi G}{3}\left(\rho + 3P\right), \tag{3.80}$$

therefore the condition we need to have is $\rho + 3P < 0$. So, since the energy density is positive, the condition is $P < -\rho/3$.

Types of inflation Another way to express the parameter $\ddot{a}$ is as the derivative of $\dot{a} = Ha$:

$$\ddot{a} = \dot{a}H + a\dot{H} = a(H^2 + \dot{H}) > 0\,, \tag{3.81}$$

so the condition can also be expressed as $H^2 + \dot{H} > 0$.

We have shown earlier that, neglecting curvature, the time-dependence of the scale factor looks like

$$a(t) = a_*\left(1 + \frac{3}{2}\left((1+w)H_*(t - t_*)\right)\right)^{2/(3(1+w))}\,, \tag{3.82}$$

and it is reasonable to use this result since, as we will discuss in this section, inflation makes curvature negligible.

We can characterize the solutions based on the sign of $\dot{H}$, which determines whether the equation of state parameter $P/\rho = w$ is larger or smaller than -1: the possibilities are

1. $\dot{H} < 0$ while $\dot{H} + H^2 > 0$: this corresponds to $-1 < w < -1/3$, which in General Relativity-speak is called a *violation of the weak energy condition*, and inserting it in our general solution gives $a(t) \propto t^\alpha$ for some $\alpha > 0$, a **power-law inflation**;
2. $\dot{H} = 0$: this is a De Sitter, dark-energy dominated universe, whose scale factor evolves like $a(t) = \exp(Ht)$, corresponding to $w = -1$;
3. $\dot{H} > 0$ (and so also $\dot{H} + H^2 > 0$), which corresponds to $w < -1$ and which gives us $a(t) \propto (t - t_{\text{bounce}})^{-\alpha}$ with $\alpha > 0$, a singularity in the future.

The boundary at $w = -1$ is called the *phantom divide*.

An estimate of the inflation e-foldings. By how much does the early universe need to inflate? The condition we need to impose is that the comoving radius of the universe at some early time, $r_H(t_0)$, should be larger than the current one, $r_H(t_0)$. Since $r_H = d_H/a$, we can write this inequality as

$$\frac{d_H(t_{\text{in}})}{a(t_{\text{in}})}a(t_f) \geq \frac{d_H(t_0)}{a(t_0)}a(t_f)\,, \tag{3.83}$$

where we multiplied both sides by the scale factor calculated at t_f, a time corresponding to the end of inflation, a minimum for the scale factor. Let us define $Z_{\text{min}} = a(t_f)/a(t_{\text{in}})$. This will be $\gg 1$, and it will describe by how much the universe inflated.

In our rough approximation $d_H \sim H^{-1}$, so we can say that the boundary of the inequality, the minimum inflationary expansion, will be

$$Z_{\text{min}} = \frac{d_H(t_0)}{d_H(t_{\text{in}})}\frac{a(t_f)}{a(t_0)} \tag{3.84}$$

$$= \frac{H(t_{\text{in}})}{H(t_0)}\frac{a(t_f)}{a(t_0)} \tag{3.85}$$

$$= \frac{H(t_{\text{in}})}{H(t_f)}\frac{H(t_f)}{H(t_0)}\frac{a(t_f)}{a(t_0)} \tag{3.86}$$

$$Z_{\text{min}}\frac{H_f}{H_{\text{in}}} = \frac{H_f}{H_0}\frac{a_f}{a_0}\,. \tag{3.87}$$

Denoting $H(t_i) \equiv H_i$ and similarly for a.

Now, we want to put some numbers into this expression: we know from the first Friedmann equation that (as long as there is no spatial curvature) $H^2 \propto \rho$, while the third tells us that $\rho \propto a^{-3(1+w)}$: therefore, $H \propto a^{-3\frac{1+w}{2}}$. Since we are working with ratios, proportionality is all we need. Now, what w should we use? At any stage in the evolution of the universe there are several fluids, but in order to simplify the calculation we will only consider the dominant one and neglect dark energy in the current phase of the evolution of the universe. In the inflationary phase we will have an undetermined $w = w_{\text{inf}}$, so that

$$\frac{H_f}{H_{\text{in}}} = \left(\frac{a_f}{a_{\text{in}}}\right)^{-3\frac{1+w}{2}}, \tag{3.88}$$

so the left-hand side of the equation reads

$$Z_{\text{min}}\frac{H_f}{H_{\text{in}}} = Z_{\text{min}}^{1-3\frac{1+w_{\text{inf}}}{2}} = Z_{\text{min}}^{\frac{-1-3w_{\text{inf}}}{2}} = Z_{\text{min}}^{\left|\frac{1+3w_{\text{inf}}}{2}\right|}, \tag{3.89}$$

since $w_{\text{inf}} > -1/3$ means $1 + 3w_{\text{inf}} > 0$. The right-hand side has precisely the same form, so we can express it as

$$\frac{H_f}{H_0}\frac{a_f}{a_0} = \left(\frac{a_f}{a_0}\right)^{-\frac{1+3w}{2}}, \tag{3.90}$$

where w is that of the dominant fluid from the end of inflation to now. The issue is, there is not a single one! In the early stages radiation was dominant, then matter started dominating (now dark energy is dominant, but we shall not worry about it). So, we split the term in two, with the radiation-matter equality being the breaking point. The earlier radiation-dominated phase is characterized by $w = 1/3$, while the latter matter-dominated phase is characterized by $w = 0$, so we can compactly write the term as

$$\frac{H_f}{H_0}\frac{a_f}{a_0} = \left(\frac{a_f}{a_{\text{eq}}}\right)^{-\frac{1+1}{2}}\left(\frac{a_{\text{eq}}}{a_0}\right)^{-\frac{1+0}{2}} = \left(\frac{a_{\text{eq}}}{a_f}\right)\left(\frac{a_0}{a_{\text{eq}}}\right)^{1/2}. \tag{3.91}$$

With this result, we now have an almost explicit expression for Z_{min}:

$$Z_{\text{min}} = \left[\left(\frac{a_{\text{eq}}}{a_f}\right)\left(\frac{a_0}{a_{\text{eq}}}\right)^{1/2}\right]^{\left|\frac{2}{1+3w_{\text{inf}}}\right|} = \left[\left(\frac{a_0}{a_f}\right)\left(\frac{a_0}{a_{\text{eq}}}\right)^{-1/2}\right]^{\left|\frac{2}{1+3w_{\text{inf}}}\right|}. \tag{3.92}$$

The ratio a_0/a_{eq} can be written as $1 + z_{\text{eq}} \approx 2.3 \times 10^4 \Omega h^2 \approx 10^4$ (very roughly). The other rough estimate we make is to apply Tolman's law, so that

$$\frac{a_0}{a_f} \approx \frac{T_f}{T_0} = \frac{T_f}{m_p} \underbrace{\frac{m_p}{T_0}}_{\sim 10^{32}};\tag{3.93}$$

properly speaking this only holds when there is radiation dominance so it does not apply for the whole range in which we are applying it (up to today), but we are estimating the order of magnitude *of an exponent*, so even an order-of-magnitude error is not an issue. We normalized by the Planck mass (or temperature, since we are using natural units) since the temperature T_f at the end of inflation probably was of that order of magnitude. The final estimate we get is

$$Z_{\text{min}} \approx \left[10^{30} \frac{T_f}{m_p} \right]^{\left| \frac{2}{1+3w_{\text{inf}}} \right|}.\tag{3.94}$$

We do not know what w_{inf} is besides it being smaller than $-1/3$; let us say that it is of the order of -1 like the current dark-energy dominated phase. If we further assume that $T_f/m_p \sim 1$,[12] we find $Z_{\text{min}} \sim 10^{30} \approx \exp(70)$.

This is often written as "70 *e*-foldings", meaning 70 *e*-fold increases in size.

3.3.2 The flatness problem

Now, let us consider the *flatness problem*, which was first proposed by Dicke and Peebles in 1986.

Cannot seem to find the paper or article...

The first Friedmann equation can be rearranged as

$$\underbrace{\frac{3a^2H^2}{8\pi G}}_{=a^2\rho_C} = \rho a^2 - \frac{3k}{8\pi G}\tag{3.95}$$

$$(\rho_C - \rho)a^2 = \left(\frac{1}{\Omega} - 1 \right) \rho a^2 = -\frac{3k}{8\pi G} = \text{const},\tag{3.96}$$

where $\rho_C = 3H^2/8\pi G$ is the critical density. The right-hand side only contains constants, therefore the left-hand side must be constant as well. As the universe evolves ρ decreases while a increases; however $\rho \propto a^{-3(1+w)}$ by the third Friedmann equation, meaning that the term scales like $\rho a^2 \propto a^{-1-3w}$: as long as $w > -1/3$, which is the case for most of the universe's evolution (with radiation and matter dominance) the term is decreasing, meaning that the other term must be increasing.

[12] The constraints on T_f are the parameters of baryogenesis (we can observe it through the isotope distribution early galaxies) and the fact we have not detected gravitational waves from this early time. These tell us that T_f/m_p cannot be larger than 10^{-3}.

If $k = 0$ we also have $\Omega \equiv 1$, so the point is moot, however as we have already mentioned there are reasons to think this is unlikely.

So, let us consider the case $k = \pm 1$: then, the term $\Omega^{-1} - 1$ must scale like a^{1+3w} in order to balance the other.

If we assume $w = 1/3$ for all times we get (using the fact that $a \propto (1+z)^{-1}$ and Tolman's law) that the term calculated at a redshift z is given by:

$$\Omega^{-1}(z) - 1 = (\Omega_0^{-1} - 1)(1+z)^{-2} = (\Omega_0^{-1} - 1)\left(\frac{T_0}{T(z)}\right)^2 . \tag{3.97}$$

The reasoning in [Pac18a] looks way more complicated, but this way seems just as valid...

The assumption $w \equiv 1/3$ is not really correct, but the result would only change by a few orders of magnitude if we did the calculation properly, and as before we are giving only a rough estimate *of an exponent*.

Let us extend this line of reasoning back to the Planck epoch, since beyond that our theory of gravity might behave differently so it is not justified to apply Friedmann's equations.

If we compute T_{Pl}/T_0 we get approximately 10^{32}. This means that

$$\Omega^{-1}(z_{\text{Pl}}) - 1 \approx (\Omega_0^{-1} - 1)10^{-64} . \tag{3.98}$$

The correct calculation, keeping track of the dominant fluid in each phase, gives 10^{-60} instead of 10^{-64}: not that significant a difference when discussing these kinds of numbers.

The current estimate for Ω_0, as discussed in section 12, is *at most* of the order of 10^{-3} away from 1 — therefore, in the Planck epoch the parameter would have needed to be different from 1 by a part in 10^{-63}.

This is a type of *fine-tuning* problem: the initial conditions seem to require an "unnatural" number like $\Omega \sim 1 \pm 10^{-63}$. Typically, a fine-tuning problem is interpreted as a signal that we should improve our theory.[13]

3.3.3 Mechanisms for inflation

Both the horizon problem and the flatness problem are addressed by the theory of inflation; up until now we have seen what inflation *does*, but we still have to discuss *how* it occurs. The way to properly describe it is through the language of Quantum Field Theory in curved spacetime, which surely cannot be introduced here; this section is meant to just give a flavor of the mechanism.

In regular Quantum Field Theory the Hamiltonian of our theory is often in the form (this example is for a real scalar field, with $a^\dagger$ and a being its creation and annihilation operator and k being the momentum):

$$H = \int d^3k \, E_k() \frac{a^\dagger a + a a^\dagger}{2} = \int d^3k \, E_k \underbrace{a^\dagger a}_{N} + \underbrace{\int d^3k \, \frac{\omega_k}{2} \left[a, a^\dagger\right]}_{\to\infty} , \tag{3.99}$$

[13] This approach has also received criticism [Hos19, sec. 3.2]: we do not know what distribution the initial conditions are drawn from, so how can we say whether a certain number is more likely (or "natural") than another?

meaning that it can be expressed in terms of an integral of the number operator times the energy, plus an integral which is constant and which diverges, since $\left[a, a^{\dagger}\right] = \delta^{(3)}(0)$ and the ground state of each harmonic oscillator in our continuum of possible values of the momentum is nonzero.

As long as we are writing our theory in flat spacetime this is not an issue: we are just adding a constant to the Hamiltonian, which does not affect the equations of motions which depend on its derivatives. In GR this is not the case: this energy *gravitates*, as we have seen any energy density does.

Is the ground-state energy just an artifact of the mathematical description of the fields? If so, we do not have a problem; unfortunately this is not the case, we can see this ground state energy through the **Casimir effect**.

If we put two metallic plates close to each other in a vacuum we can detect an attractive force between them, due to the fact that long-wavelength fluctuations do not "fit" in the gap between the plates, decreasing the energy density between the plates compared to the one outside, which is equivalent to the binding energy of an attractive force.[14]

An interesting digression, but I'm not sure whether it fits in this section.

In QFT fields are classified based on how they transform under rotations: scalars do not change, vectors come back to themselves after a rotation of 2π, spinors come back to themselves after a rotation of 4π. In the standard model generally "matter" particles are spinors, while "force" particles are vectors. Scalars are rare: the only one is the Higgs boson.

We want our mechanism for inflation to be a field with a nonzero expectation value, which should also satisfy the symmetries of the FLRW metric. Our only option then is a scalar field which is homogeneous (and automatically isotropic since it does not define a direction), but which can change in time.

A vector and a spinor both define a direction in space, and thus do not satisfy the requirement of isotropy. A term like $\overline{\psi}\psi$, where ψ is a spinor and $\overline{\psi}$ is its adjoint, can actually respect the required symmetries. We will not explore this further, but it gives rise to what is called a *fermion condensate*.

So, we will add a scalar field Φ to our model.

The Lagrangian formulation of GR Usually, an action for the Standard Model particles in a general-relativistic setting will have a term containing the derivatives of the metric, S_g, and a term containing the actions of all the Standard Model particles, S_{SM}:

$$S = S_g + S_{\text{SM}} . \tag{3.100}$$

As in classical mechanics, the equations of motion are derived from a variational principle, $\delta S = 0$; however since S is a function of many fields this actually contains the equations of motion for each of them. Varying with respect to the SM fields yields their equations of

[14] This was demonstrated to actually occur: the first group to do the experiment with the original parallel-plate configuration was in Padua [Bre+02]!

motion, while varying with respect to the inverse metric $g^{\mu\nu}$ yields the Einstein equations:

$$\underbrace{R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu}}_{\propto \frac{\delta S_g}{\delta g^{\mu\nu}}} = \underbrace{8\pi G T_{\mu\nu}}_{\propto \frac{\delta S_{\mathrm{SM}}}{\delta g^{\mu\nu}}} . \tag{3.101}$$

In this Lagrangian approach, the very *definition* of the stress-energy tensor is as a certain multiple of the functional derivative of the action of the particles with respect to the inverse metric.

Coupling between a field and gravity We update the action by adding a term for the field Φ:

$$S = S_\Phi + S_g + S_{\mathrm{SM}} . \tag{3.102}$$

As usual the action is found by integrating a Lagrangian density, but since we have a nontrivial metric we need to use the invariant volume element $\mathrm{d}^4x\,\sqrt{-g}$ (which for FLRW is just $\mathrm{d}^4x\, a^3$):

$$S = \int \mathrm{d}^4x\, \sqrt{-g} \mathscr{L} . \tag{3.103}$$

The gravitational Lagrangian is given in terms of the Ricci scalar, $\mathscr{L}_g = R/16\pi G$, and varying it with respect to the metric yields the left-hand side of the Einstein equations.

A typical Lagrangian for a particle of mass m whose position is described by the coordinates q is given by a kinetic term and a potential term:

$$\mathscr{L} = \frac{m}{2} \dot{q}^2 - V(q) , \tag{3.104}$$

for a scalar field in Minkowski spacetime its equivalent would be

$$\mathscr{L}_\Phi = \frac{1}{2} \Big(\partial_\mu \Phi \Big) (\partial^\mu \Phi) - V(\Phi) . \tag{3.105}$$

In the GR case it is customary to be more explicit about the metric appearing in the kinetic term; also, the derivatives should in general become covariant ones, although in this case there is no change since the covariant derivative of a scalar is equal to its partial one: $\nabla_\mu \Phi = \partial_\mu \Phi$. The Lagrangian then becomes:

$$\mathscr{L} = \frac{1}{2} g^{\mu\nu} \partial_\mu \Phi \partial_\nu \Phi - V(\Phi) . \tag{3.106}$$

The potential may include different terms, a common one is a mass term, which looks like $V(\Phi) = m\Phi^2/2$.

If we add a massive term, proportional to $R\Phi^2$, we get that adding it to the global action looks like gravity.

Clarify

Actions are dimensionless since $\hbar = 1$, and since $\mathrm{d}s^2 = g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu$ the metric $g_{\mu\nu}$ is also dimensionless. The Riemann tensor is given by the second derivatives of the metric, so its dimension is a length to the -2, or a mass squared.

So, the dimensional analysis of $\int \mathrm{d}^4x\,\sqrt{-g}\mathscr{L}$ gives us that $\mathscr{L}$ must have dimensions of a length to the -4, or a mass to the 4. The field Φ has the dimensions of a mass, which is an inverse length. The coupling constants are conventionally taken to be dimensionless: therefore if we are to add a term to the Lagrangian, it must be $\xi\Phi^2$ times an inverse square length, which is often m^2, m being the mass of the field, while ξ is a real number.

With all of this said, in terms of dimensionality we can add to our Lagrangian a term $\xi R\Phi^2$, where R is the Ricci scalar. This is a prototype for modified GR theories.

The value of ξ is undetermined: setting $\xi = 1/6$ gives us conformal symmetry, while in other cases it is useful to set it as $\xi = 1/4$. A Weyl transformation (a local rescaling of the metric) allows us [FGN98] to remove this term: we move from the Jordan frame (where we *do* have coupling between our scalar field and the curvature, with a term such as the one we described) to the Einstein frame, in which we do not have this term, but we *do* have an additional matter-like term in the Einstein equations, a new component in the stress-energy tensor, which will look like:

$$T_{\mu\nu}(\Phi) = \Phi_{,\mu}\Phi_{,\nu} - g_{\mu\nu}\left(\frac{1}{2}g^{\rho\sigma}\Phi_{,\rho}\Phi_{,\sigma} - V(\Phi)\right), \tag{3.107}$$

where commas denote partial derivatives: $\Phi_{,\alpha} = \partial_\alpha\Phi$.

We can get an explicit solution for the solution of the equations of motion of this field by using the symmetries of our spacetime: we assume that, because of homogeneity, $\Phi(x^\mu) = \varphi(t)$.

There is another issue: in QFT any field Φ is an operator acting on a Fock space, while the left-hand side of the Einstein equation is a simple tensor — we are not quantizing space! The solution to this problem is a semiclassical mean-field approximation, which is similar to the Hartree-Fock mean-field method: we assume we are "close" to the ground state, and so we substitute the stress-energy tensor on the right-hand side with its mean value computed in the ground state of the theory:

$$G_{\mu\nu} = 8\pi G\left\langle \hat{T}_{\mu\nu}\right\rangle_0, \tag{3.108}$$

where we define the ground state $|0\rangle$ as that one with the most symmetry allowed (that is, it should be invariant under rotations and translations, the symmetries of the FLRW metric).

If we perturb the state of thee field Φ, we get $\Phi = \varphi + \delta\Phi$: so $\left\langle \Phi^2\right\rangle = \varphi^2 + 2\left\langle \varphi\delta\Phi\right\rangle + \left\langle \delta\Phi^2\right\rangle$, but the second term is zero since $\langle\delta\Phi\rangle = 0$ and φ is constant. The last term in diverges. We do not know how to deal with it. We therefore assume that it is small.

What? is this about renormalization?

When computing the stress energy tensor we get only diagonal terms: this scalar acts like a perfect fluid!

The energy density is equal to the Hamiltonian:

$$\rho = T_{00} = \frac{1}{2}\dot{\varphi}^2 + V(\varphi) = H, \tag{3.109}$$

while the pressure is the Lagrangian:

$$P = \frac{1}{2}\dot{\varphi}^2 - V(\varphi) = \mathscr{L}\,. \tag{3.110}$$

We assume that anything in the universe which is not our field behaves like radiation, with energy density ρ_r. Then, the Friedmann equations (assuming zero spatial curvature) will read

$$H^2 = \frac{8\pi G}{3}\left(\frac{1}{2}\dot{\varphi}^2 + V + \rho_r\right) \tag{3.111a}$$

$$\frac{\ddot{a}}{a} = -\frac{8\pi G}{3}\left(\dot{\varphi}^2 - V + \rho_r\right) \tag{3.111b}$$

$P_r = \rho_r/3$.

$$\dot{\rho}_{\text{tot}} = -3\frac{\dot{a}}{a}\left(\rho_{\text{tot}} + P_{\text{tot}}\right), \tag{3.111c}$$

but in the continuity equation we can split the contributions by inserting an unknown factor Γ, the transfer of energy between the field and radiation, so that the respective energy densities evolve as

$$\dot{\rho}_\varphi = -3\frac{\dot{a}}{a}\dot{\varphi}^2 + \Gamma \tag{3.112a}$$

$$\dot{\rho}_r = -4\frac{\dot{a}}{a}\rho_r - \Gamma\,. \tag{3.112b}$$

In order to see what the evolution of the field looks like we drop Γ, assuming that there is little energy transfer between matter and radiation. The equation of motion of the field reads

$$\ddot{\varphi} + 3H\dot{\varphi} = -V'\,, \tag{3.113}$$

where $V' = \partial_\varphi V$.[15]

We have already mentioned that this field behaves like a fluid: so, what is its equation of state? the definition of w is

$$w = \frac{P}{\rho} = \frac{\frac{1}{2}\dot{\varphi}^2 - V}{\frac{1}{2}\dot{\varphi}^2 + V}\,, \tag{3.115}$$

whose limiting case, when $\dot{\varphi}^2 \ll 2|V|$, is $w = -1$. As we have seen, this corresponds to an evolution of the universe with $a \sim \exp(Ht)$: an exponential expansion!

The "old inflation model" has inflation being caused by the breaking of a certain symmetry through quantum tunneling, while a new inflation model involves "slow rolling".

The equation $\ddot{\varphi} + 3H\dot{\varphi} = -V'$ looks like a regular equation of motion with a kinetic and friction term: after a time $1/H$ the "friction" velocity-dependent term dominates.

[15] This may look peculiar, but it is in fact just the Klein-Gordon equation written in curved spacetime: the difference comes about because the Dalambertian operator now reads [NS19]

$$\Box\varphi = \frac{1}{\sqrt{-g}}\partial_\mu\left(\sqrt{-g}\partial^\mu\varphi\right) = \partial_\mu\partial^\mu\phi + \frac{\partial_\mu\sqrt{-g}}{\sqrt{-g}}\left(\partial^\mu\varphi\right) = \ddot{\varphi} + \dot{\varphi}\frac{\partial_t(a^3)}{a^3} = \ddot{\varphi} + 3\dot{\varphi}\frac{\dot{a}}{a}\,. \tag{3.114}$$

Then we get a slow-roll (friction-dominated) regime: $H^2 = \frac{8\pi G}{3} V$ and $\dot{\varphi} \approx -V'/3H$.

Mention of chaotic inflation and many more things... could not really form a coherent narrative. Should try again after reading [LC02, sec. 7.11].

This section could probably do with some pictures of potentials...

Let us consider a region of radius approximately $1/H(t_b)$, where t_b is a certain moment before the start of the inflationary phase. This region is expanded by many e-foldings through inflation, such that any inhomogeneities are smoothed out: this is the **cosmic no-hair theorem** [LC02, pag. 159].

So, there might have been perturbations before inflation: we cannot know. Perturbations on scales larger than the cosmological horizon are not perceivable as perturbations: we only perceive our local mean value.

Reheating The energy density of radiation through the inflationary period scales as $\rho_r \propto a^{-4} \propto e^{-4Ht}$, while the one of matter instead it scales as $\rho_m \propto a^{-3} \propto e^{-3Ht}$ since $a \propto e^{Ht}$. Qualitatively speaking, as the universe rapidly expands it becomes basically *devoid of particles*, and its temperature drops substantially.

At the end of inflation, *reheating* takes place, substantially raising the temperature and allowing for the formation of most of the SM particles we observe today. This is due to the latent energy released by our scalar field due to its coupling to the rest of the universe, which acts as a sort of viscous force. Intuitively speaking, the field "falls" into its ground state and then oscillates ever slower, dissipating its energy in the process [LC02, fig. 7.6].

3.4 Baryogenesis

This is the process which led to the formation of baryons.

The main issue we seek to address is that of **baryon asymmetry**. In the Standard Model each fermion has a corresponding antifermion, and the same holds for the composite particles of quarks, such as mesons and baryons.

We usually talk of baryons only; they are characterized by the conserved baryon number B, which is the number of baryons minus the number of antibaryons.[16]

Matter and antimatter annihilate if they meet, producing radiation (γ photons and/or other high-energy particles). This allows us to investigate the presence of antimatter in the universe: if there were patches of antimatter as well as patches of matter their boundaries would be regions of great γ-ray emission, which we would be able to detect as a background. Observations then allow us to rule out the presence of antimatter regions large enough to have $B = 0$ in the observable universe — small patches of antimatter may exist, but there are not enough of them to balance out all the matter [CDG98].

This constitutes the problem of baryon asymmetry: it seems unnatural for the universe to start off with more baryons than antibaryons, all the known Standard Model interactions conserve B, yet we observe more baryons than antibaryons.

[16] This number can actually be computed from the number of quarks already, so it makes sense to discuss it even before the formation of proper baryons.

In particle physics the absolute number B is used, but in cosmology absolute numbers are not used: we work in terms of densities. All number densities scale like a^{-3}, so we need to normalize the difference $n_b - n_a$ with another number density: we define: $\eta/2 = (n_b - n_a)/(n_b + n_a)$. Here b means baryons, while a means antibaryons.

The reason for the factor 2 is given by how we can actually estimate this parameter: in the early universe, when the baryons were still relativistic, processes like $b + \overline{b} \leftrightarrow 2\gamma$ occurred at thermal equilibrium, so roughly speaking we should have had $n_b \approx n_a \approx 2n_\gamma$.

What would be the proper chemical-equilibrium considerations?

Now, all of these densities scaled like a^{-3} from the moment the baryons became nonrelativistic to now. So, we have

$$2\frac{n_b - n_a}{\underbrace{n_b + n_a}_{2n_\gamma}} \approx \frac{n_{0b} - \overbrace{n_{0a}}^{\approx 0}}{n_{0\gamma}} \approx \frac{n_{0b}}{n_{0\gamma}} = \eta_0 \,. \tag{3.116}$$

This allows us to estimate the asymmetry constant η with parameters measured today: the baryon number density and the photon number density.

For the baryon number density we can estimate [17]

$$n_{0b} \approx \frac{\rho_{0b}}{m_p} = \frac{\Omega_{0b}\rho_C}{m_p} \approx 5.46 \times 10^{-7}\,\mathrm{cm}^{-3} h^2 \,, \tag{3.117}$$

$\Omega_{0b} \approx 0.0486.$

while for the photons we can find the number density by integrating the CMB Planckian: the result is [18]

$$n_{0\gamma} \approx \frac{2\zeta(3)}{\pi^2} T_{0\gamma}^3 \left(\frac{k_B}{\hbar c}\right)^3 \approx 410\,\mathrm{cm}^{-3} \,. \tag{3.118}$$

[17] The result comes about from the following code:

```python
from astropy.cosmology import Planck15 as cosmo
import numpy as np
import astropy.units as u
from astropy.constants import codata2018 as ac
H0 = u.littleh *100 * u.km/u.s / u.Mpc
rhoC = 3 * H0**2 / (8 * np.pi * ac.G)
(rhoC * cosmo.Ob0 / ac.m_p).to(u.cm**-3 * u.littleh**2)
```

[18] The result comes about from the following code:

```python
from astropy.cosmology import Planck15 as cosmo
import numpy as np
import astropy.units as u
from astropy.constants import codata2018 as ac
z3 = np.sum([n**-3 for n in range(1, 1000000)])
(2 * z3 / np.pi**2 * cosmo.Tcmb0**3 * (ac.k_B / ac.hbar / ac.c)**3).cgs
```

Combining these results yields

$$\eta_0 \approx 3 \times 10^{-8} \Omega_{0b} h^2 \approx 1.3 \times 10^{-9} h^2 \approx 6.1 \times 10^{-10} \,. \tag{3.119}$$

How do we interpret this result? It means that in the radiation-dominated epoch the asymmetry between baryons and antibaryons was very slight, about one part in a billion; most of the pairs annihilated but a bit of matter was leftover, and that is the matter we currently have.

In 1966, the Soviet physicist Sakharov postulated that in order to generate baryon-antibaryon asymmetry in the early universe there are three necessary conditions:

1. violation of baryon number conservation;
2. C and CP violation (while CPT symmetry must hold for any well-behaved QFT);
3. out-of-equilibrium processes.

B violation is needed for the existence of processes which can generate more baryons than antibaryons. C violation is needed because otherwise any process which generates baryons would have a counterpart generating antibaryons occurring with the same probability. CP violation is needed because otherwise processes generating left-handed baryons would be balanced by processes generating right-handed antibaryons, and vice versa. Out-of-equilibrium processes are needed because otherwise B-increasing and B-reducing processes would balance out.

3.5 Decoupling of particle species

Earlier we made the claim that neutrinos decouple at $T \approx 1\,\text{MeV}$: let us justify this, and explore the concept of a particle species being **decoupled** in general.

We define the *interaction rate* Γ: it is the number of interactions per unit time that each particle undergoes.

The Hubble rate $H = \dot{a}/a$ is also, dimensionally, an inverse time, so we can compare H and Γ directly. What does this comparison tell us? We have seen earlier that the age of the universe is roughly given by $1/H$, while $1/\Gamma$ is the average time for an interaction to occur (modelling the interaction times as a Poisson process). So, if $1/H < 1/\Gamma$, then on average the particle species has undergone on average *less than one* interaction in the whole existence of the universe. This condition is called decoupling, and is equivalent to $\Gamma < H$. If $\Gamma \ll H$ then the interaction basically does not happen.

In general, Γ can be calculated as $\Gamma = n \langle \sigma v \rangle$, where n is the number density, v is the velocity of the particles, σ is the cross section of the interaction, and we average over the velocity distribution of the particles.

In the Standard Model, interactions are mediated by gauge bosons, which can be either massless (like the photon) or massive (like the weak interaction $W^\pm$ and Z bosons[19]).

[19] Properly speaking, these bosons are not massive in general since they acquire their mass through the spontaneous breaking of the $SU(2)_L \times U(1)_Y$ symmetry of the Standard Model to $U(1)_{\text{em}}$, which happens

For massless boson mediators, the cross-section is (roughly) given by $\sigma \sim \alpha^2/T^2$, where $g = \sqrt{4\pi\alpha}$.

For massive boson mediators with mass m_x, we need to distinguish between two cases: for temperatures $T \leq m_x$, the cross section is of the order $\sigma \sim G_x^2 T^2$, while for higher temperatures $T > m_x$ we have $\sigma \sim \alpha^2/T^2$ as in the massless case.[20] Typically, $G_x = \alpha/m_x^2$.

Let us then estimate the decoupling temperatures in the two cases. In our computation, v is the relative velocity between the two particles; in both cases we assume that the particles are relativistic, therefore $v \approx 1$. We will make very rough estimates, neglecting multiples like π in front of our expressions; the things we want to get right are the dimensionality and the asymptotic scaling of the equations.

Massless boson decoupling We know that the number density of particles scales like $n \sim T^3 \sim a^{-3}$ in this radiation-dominated epoch, therefore the interaction rate will scale like

$$\Gamma \sim T^3 \sigma \sim T^3 \frac{\alpha^2}{T^2} \sim \alpha^2 T\,. \tag{3.121}$$

What is the scaling of H? We can recover it from the first Friedmann equation (see equation (3.63)):

$$H = \sqrt{\frac{8\pi G}{3} g_*^{1/2} \left(\frac{\pi^2}{30}\right)^{1/2}} T^2 \sim \frac{T^2}{m_{\text{pl}}}\,, \tag{3.122}$$

since $G \sim 1/m_{\text{pl}}^2$. The numeric factor we are neglecting is about a factor 10, only one order of magnitude, which is fine given the roughness of our calculation.

Let us then see when $\Gamma/H < 1$, which corresponds to decoupling:

$$\frac{\Gamma}{H} \sim \frac{\alpha^2 T m_{\text{pl}}}{T^2} \sim \alpha^2 m_{\text{pl}} \frac{1}{T}\,, \tag{3.123}$$

so we have decoupling when $T > \alpha^2 m_{\text{pl}}$.

For electromagnetism $\alpha \approx 1/137$, so at temperatures larger than $T \approx 10^{16}\,\text{GeV}$ this massless photon is decoupled.

Note that decoupling occurs for *very high* temperatures, meaning very early times, close to the Planck epoch; further, the photons start off decoupled and then couple. After this, there is no lower bound: they may remain coupled for arbitrarily low temperatures.

---

below the scale of electroweak symmetry breaking $\sim 10^2\,\text{GeV}$. The bosons being massive yields an exponential cutoff e^{-mr} in the expression of the potential for the interaction (this is the *Yukawa potential*), so at energies higher than the electroweak SB scale the weak interaction becomes long-range.

[20] The reason for this is that when we compute the cross section for an interaction with a massive boson we find an expression like

$$\sigma \sim \frac{s}{(s - m_x^2)^2}\,, \tag{3.120}$$

where s is the square of the center-of mass energy of the interaction; if there is thermal equilibrium then typically we will have $s \sim T^2$. We can then see that at high energies the expression is asymptotically $\sigma \sim T^{-2}$, while at low energies it is $\sigma \sim T^2/m_x^4$. This expression also holds for $m_x = 0$.

Some comments are made about gravitational interactions, but what would α_{grav} be?

The calculation seems to yield $T \approx 10^{15}\,\text{GeV}$, or even less if we account for g_*... I guess it does not really matter but it seemed curious.

Massive boson decoupling It is of interest for us to consider the *low-energy* situation, in which $T < m_x$. The opposite limit, $T \gg m_x$ behaves like the massless case: and we find an upper limit above which there is decoupling.

So, we have $\Gamma \sim T^3 G_x^2 T^2 = G_x^2 T^5$, to compare with $H \sim T^2/m_{\text{pl}}$: as before we take the ratio, to find

$$\frac{\Gamma}{H} \sim \frac{T^5 G_x^2}{T^2/m_{\text{pl}}} \sim G_x^2 m_{\text{pl}} T^3 \,, \tag{3.124}$$

so we have decoupling ($\Gamma/H < 1$) when $T < m_{\text{pl}}^{-1/3} G_x^{-2/3}$. This is a *lower bound*! The situation is qualitatively different from the massless case.

For the weak interaction, G_x is called the Fermi constant $G_F \sim (300\,\text{GeV})^{-2}$:[21] this yields the bound

$$T \lesssim 1\,\text{MeV} \,, \tag{3.125}$$

which is why below 1 MeV neutrinos are decoupled.

Then Pacciani [Pac18a] applies the formula to gravitation setting $G_x = G$ to find that it decouples at $T \sim m_P$, but the graviton is massless!

To summarize, interactions which are mediated by massless particles couple at a certain (high) temperature and are thereafter coupled. On the other hand, interactions which are mediated by massive particles couple at a certain (high) temperature, stay coupled for a while, and then decouple again at a relatively low temperature.

3.6 Hydrogen recombination

In the early universe, at $z \sim$ a few $\times 1000$, there were free electrons and free protons. The temperature had dropped below the binding energy of hydrogen, 13.6 eV, quite a long time earlier, around $z \sim 5 \times 10^4$, however as we will see in more detail the conditions were such that most hydrogen was still ionized.

Around $z \sim 2000$ the process of recombination[22] started. Then, around $z \sim 1400 \div 1600$ [LC02, table 9.1] the process started really picking up, crossing the half-way point for the fraction of hydrogen which was ionized.

[21] This value can be experimentally measured from weak interactions, but it is consistent with the masses of the mediators being of the order of 100 GeV.

[22] The process $e^- + p \to H + \gamma$ is known as "recombination" for historical reasons, although electrons and protons were not re-combining, since bound hydrogen could not have existed earlier in the history of the universe.

At around $z \sim 1100$ the fraction of ionized hydrogen became so low that the universe became transparent to radiation. Photons which were being constantly scattered reached their last scattering and then kept going; the temperature at that point was $T \sim 3000\,\mathrm{K}$, so now we see the CMB as a thermal distribution at a temperature of $3000\,\mathrm{K}/(1+z) \approx 2.7\,\mathrm{K}$.

Now, let us get to the details of how this all happened.

In order to precisely treat the process $e^- + p \leftrightarrow H + \gamma$ we would need all the scattering matrix elements, as well as all the phase space densities of the particles. This would allow us to treat a general distribution in phase space, even out of equilibrium.

What we will do instead is to provide a bulk estimate, making use of the Saha equation, which assumes we are at both thermal and chemical equilibrium. It relates the chemical potentials of the particles as $\mu_e + \mu_p = \mu_H + \mu_\gamma$; however we know that $\mu_\gamma = 0$, so we have

$$\mu_e + \mu_p = \mu_H \,. \tag{3.126}$$

Both electrons, protons and hydrogen atoms were nonrelativistic at this stage: therefore their number densities can be approximated with Boltzmann statistics (the low-energy approximation: equation (3.51)),

$$n_e = g_e \left(\frac{m_e T}{2\pi} \right)^{3/2} \exp\left(\frac{\mu_e - m_e}{T} \right) \tag{3.127}$$

$$n_p = g_p \left(\frac{m_p T}{2\pi} \right)^{3/2} \exp\left(\frac{\mu_p - m_p}{T} \right) \tag{3.128}$$

$$n_H = g_H \left(\frac{m_H T}{2\pi} \right)^{3/2} \exp\left(\frac{\mu_H - m_H}{T} \right) . \tag{3.129}$$

The statistical weights here are $g_e = g_p = 2$ and $g_H = 4$ [LC02, pag. 194].[23] Also, we will later need the fact that the universe being globally neutral implies $n_e = n_p$.

Let us now introduce the total baryon density, n_b. In principle, we should account for Helium: as we will see, a couple of minutes after the Big Bang He-4 nuclei started to form, but they made up only something like 25% of the mass, which means 6% of the number density: so we ignore them and say that the number density of baryons is

$$n_b = n_p + n_H \,. \tag{3.130}$$

The quantity we want to describe is how much hydrogen is still ionized: this is given by the ratio of the free electrons to the total baryons, n_e/n_b. This is called the **ionization fraction** X_e, and is also equal to n_p/n_b since the universe must be globally neutral.

We expect to have $X_e = 1$ in the early universe, and $X_e = 0$ after the end of reionization.

This is a good model to keep in mind although it is not precisely correct, since simulations show that there will be some *residual ionization*: X_e only goes to around $10^{-4} \div 10^{-5}$ at the end of reionization. Also, in the modern universe a large fraction of the hydrogen has become ionized again, especially in the intergalactic medium; this is likely due to the energy injected into it by structure formation, which definitely breaks the approximation of

[23] These numbers are given as fact here, they come from a quantum-mechanical study of the system.

global thermal and chemical equilibrium. This is not a concern for us: here, we only wish to describe the processes in the interval $1000 \lesssim z \lesssim 2000$.

The binding energy of the hydrogen is $B = m_p + m_e - m_H = 13.6\,\mathrm{eV}$, so instead of m_H we can write $m_H = m_p + m_e - B$. Let us insert this expression and the Saha equation in the expression for the hydrogen number density; then we will recognize part of the expressions for the proton and electron number densities, which we can

$$
\begin{aligned}
n_H &= g_H \left(\frac{m_H T}{2\pi}\right)^{3/2} \exp\left(\frac{\mu_H - m_H}{T}\right) && (3.131a)\\
&= g_H \left(\frac{m_H T}{2\pi}\right)^{3/2} \exp\left(\frac{\mu_e + \mu_p - m_e - m_p + B}{T}\right) && (3.131b)\\
&= \underbrace{\frac{g_H}{g_e g_p}}_{=1} \cancel{\left(\frac{m_H T}{2\pi}\right)^{3/2}} \left(\frac{m_e T}{2\pi}\right)^{-3/2} \cancel{\left(\frac{m_p T}{2\pi}\right)^{-3/2}} n_e n_p \exp\left(\frac{B}{T}\right) && (3.131c)\\
\frac{n_H}{n_e n_p} &= \left(\frac{m_e T}{2\pi}\right)^{-3/2} \exp\left(\frac{B}{T}\right) && (3.131d)\\
\frac{n_b - n_p}{n_p^2} &= \left(\frac{m_e T}{2\pi}\right)^{-3/2} \exp\left(\frac{B}{T}\right), && (3.131e)
\end{aligned}
$$

which we can manipulate, using the following identity:

$$
\frac{n_b - n_p}{n_p^2} = \frac{n_b \left(1 - n_p / n_b\right)}{n_b^2 X_e^2} = \frac{1}{n_b} \frac{1 - X_e}{X_e^2}, \tag{3.132}
$$

where we use $n_e = n_p$ and the definition of $X_e = n_p / n_b$. Then, we bring the n_b to the other side of the equation and multiply and divide by the photon number density, which is given by

$$
n_\gamma = \frac{2\zeta(3) T^3}{\pi^2}. \tag{3.133}
$$

With this, we can insert the baryon fraction η_b (which is conserved, so we can use its current value):

$$
\begin{aligned}
\frac{1 - X_e}{X_e^2} &= \underbrace{\frac{n_b}{n_\gamma}}_{\eta_p} n_\gamma \left(\frac{m_e T}{2\pi}\right)^{-3/2} \exp\left(\frac{B}{T}\right) && (3.134a)\\
&= \eta_0 \frac{2\zeta(3) T^3}{\pi^2} \left(\frac{m_e T}{2\pi}\right)^{-3/2} \exp\left(\frac{B}{T}\right) && (3.134b)\\
&= \eta_0 \frac{4\sqrt{2}\zeta(3)}{\sqrt{\pi}} \left(\frac{T}{m_e}\right)^{3/2} \exp\left(\frac{B}{T}\right), && (3.134c)
\end{aligned}
$$

which we can solve numerically to find X_e as a function of temperature, or of redshift. The results are shown in figure 3.1.

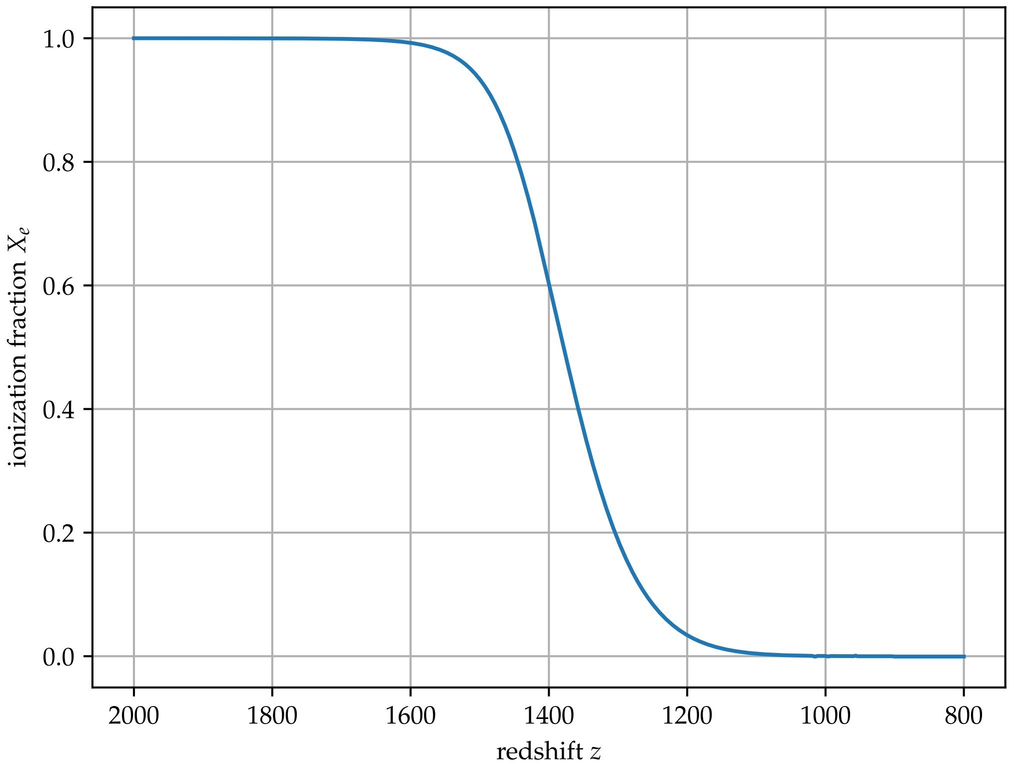

Figure 3.1: Ionization fraction as a function of redshift: a numerical solution of equation (3.134c).

Let us try to understand what is going on. Roughly speaking, we get intermediate values for X_e, like 0.5, when the right-hand side is of order 1. If the right-hand side only contained the terms $(T/m_e)^{3/2} \exp(B/T)$ without any multiplicative factor in front (meaning, roughly, that we had $\eta_0 \sim 1$), then we would see recombination starting to occur already at $z \sim 4000$ ($T \sim 1\,\text{eV}$), and by $z \sim 2000$ if would almost be over. The reason this does not happen is that η_0 is *very small*: there are a lot of photons, many more than the baryons, which are ready to dissociate any atoms which start to form.

So, hydrogen is truly formed only when $T \sim 0.3\,\text{eV}$, much lower than its ionization energy.

But even with $\eta_0 \sim 1$ we get formation at $T \sim 1\,\text{eV}$, an order of magnitude less than B! is there an intuitive argument as to why this is the case?

The estimates we gave do depend on the value we assign to Ω_{0b} and h, however within the currently accepted experimental ranges the main predictions do not vary substantially.

The process of recombination is gradual, but we can choose a conventional redshift as, for example, the time when we reach an arbitrary threshold like $X_e = 0.1$.

The interaction rate between photons and free electrons is $\Gamma_\gamma = n_e \sigma_T c$, where σ_T is the Thompson cross section while n_e is the number density of free electrons, which can be estimated as $n_e = n_b X_e \approx \rho_C \Omega_b X_e / m_p$. Then, we can estimate the moment of the last scattering by checking the decoupling condition: $\Gamma_\gamma < H$. With the numbers given earlier, we find $z \sim 1120$: a very good approximation to the currently accepted value $z \sim 1089$!

Then Pacciani has an argument as to why $T \sim a^{-2}$ for matter, should this be included here?

3.7 Primordial nucleosynthesis

Already in the 1940s it was noticed by Alpher, Bethe as well as Gamow that the abundances of certain nuclides could not be explained if they were formed in stellar interiors alone. Specifically, the issue is with the abundances of light elements: deuterium ^{2}H, as well as ^{3}He and ^{4}He, while heavier nuclides (with mass number $A \geq 7$) could not form in the early universe [LC02, sec. 8.6.1].[24]

The most abundant of these light nuclides by far is ^{4}He, whose mass fraction is denoted as $Y \approx 25\,\%$, while its number fraction is approximately $6\,\%$. Our model will need to yield this many helium-4 nuclides after primordial nucleosynthesis.

A mechanism for the synthesis of these light nuclides in the early universe is needed.

We shall model this mechanism, under the following assumptions (which are, as far as we know, roughly verified):

1. the universe passed through a very high temperature phase, with $T > 10^{12}\,$K, during which thermal equilibrium held;
2. the universe at this stage is described by General Relativity and the Standard Model of particle physics, and it is homogeneous and isotropic;
3. the chemical potentials for the neutrinos μ_ν have certain upper bounds, such that the number of neutrino types is approximately 3;
4. there is no matter-antimatter separation (as in, antimattr "bubbles");
5. there are no strong magnetic fields;
6. the number of exotic particles has a certain upper bound (it is small compared to the number of photons).

The main formation channels in the early universe are:

1. $n + p \leftrightarrow d + \gamma$ ($d = {}^2$H denotes deuterium);

[24] Environments which allow their production (stellar interiors) have lower temperature but a higher density, since the main obstruction to their production is the absence of stable nuclides at $A = 5$ and $A = 8$ — the environment needs to raise the odds of an unstable ^{8}Be nuclide colliding with an α particle to form carbon before it decays.

2. $d + d \leftrightarrow {}^3\mathrm{He} + n$;

3. ${}^3\mathrm{He} + d \leftrightarrow {}^4\mathrm{He} + p$.

Note that these processes, unlike the stellar ones, do not involve the weak interaction: neutrons and protons do not turn into each other. In stars, there are no free neutrons so this process is not possible.

The slowest process of the three is the first, since it is heavily affected by photons, which destroy deuterium. After we have produced deuterium, Helium-4 is readily produced.

In order to find out how much deuterium we have, we need the proton-to-neutron ratio. We are working at energies of around 1 MeV, so protons and neutrons are not relativistic anymore; *as long as they are in equilibrium* through weak processes both will obey Boltzmann statistics, so for $i = n, p$:

$$n_i = g_i \left(\frac{m_i T}{2\pi} \right)^{3/2} \exp\left(\frac{\mu_i - m_i}{T} \right) , \tag{3.135}$$

meaning that their number ratio is given by[25]

The explanation in [LC02] is not the same as the one given by Pacciani [Pac18b]. . . I'm inclined to trust the former.

$$\frac{n_n}{n_p} \sim \exp\left(- \frac{m_n - m_p}{T} \right) , \tag{3.136}$$

where $m_n - m_p \approx 1.3\,\mathrm{MeV} \approx 1.5 \times 10^{10}\,\mathrm{K}$.

The proton is the lightest baryon and is therefore stable; while the neutron is unstable: it can decay through the weak-interaction processes

1. $n + \nu_e \leftrightarrow p + e^-$;

2. $n + e^+ \leftrightarrow p + \overline{\nu_e}$;

3. $n \to p + e^- + \overline{\nu_e}$.

The neutron fraction keeps decreasing as T decreases and these processes keep happening, however as we have previously discussed at around $T_{d\nu} \sim 1\,\mathrm{MeV}$ neutrinos decouple, at which point the first two back-and-forth reactions stop, and we are left with $n_n / n_p \approx \exp(-\Delta m / T_{d\nu}) \approx 0.27$.

We define the number fraction of neutrons, which is approximately

$$X_n(t) \equiv \frac{n_n}{n_n + n_p} \approx 0.21\,. \tag{3.137}$$

[25] We are neglecting the chemical potentials since, as explained by Lucchin and Coles [LC02, sec. 8.6.2], as long as both weak and electromagnetic interactions are in chemical equilibrium the chemical potentials are forced to be zero by all the balance equations.

The third reaction, which is β^- decay, keeps occurring since it does not require the presence of neutrinos, so after the decoupling of neutrinos the number fraction decays exponentially as:

$$X_n(t) = X_n(t_{d_\nu}) \exp\left(-\frac{t - t_{d_\nu}}{\tau_n}\right), \tag{3.138}$$

where $\tau_n = \log 2\tau_{1/2}$, and the half-life of neutrons is given by $\tau_{1/2} \approx (10.5 \pm 0.2)$ min. So, each minute neutrons stay unbound some of them are decaying; the process of deuterium is however rather fast as we shall see, so that not many of them are lost.

Let us then move to deuterium formation: its binding energy is around $B_d = m_p + m_n - m_d \approx 2.2$ MeV. We proceed exactly like we did with hydrogen: since $\mu_p + \mu_n = \mu_d$, the deuterium number density can be expressed as

$$n_d = g_d \left(\frac{m_d T}{2\pi}\right)^{3/2} \exp\left(\frac{\mu_d - m_d}{T}\right) \tag{3.139}$$

$$= \frac{g_d}{g_p g_n} n_n n_p \left(\frac{m_d}{m_n m_p}\right)^{3/2} \left(\frac{T}{2\pi}\right)^{-3/2} \exp\left(\frac{B_d}{T}\right), \tag{3.140}$$

which, dividing through by n_b and using $g_d = 3$ (since deuterium has spin 1) and $g_p = g_n = 2$, can be expressed as:

$$X_d = \frac{3}{4} n_b X_n X_p \left(\frac{m_d}{m_n m_p}\right)^{3/2} \left(\frac{T}{2\pi}\right)^{-3/2} \exp\left(\frac{B_d}{T}\right) \tag{3.141a}$$

$$= \frac{3}{4} \eta_0 X_n X_p \left(\frac{m_d}{m_n m_p}\right)^{3/2} \frac{2\zeta(3)}{\pi^2} (2\pi T)^{3/2} \exp\left(\frac{B_d}{T}\right) \tag{3.141b}$$

Substituted $n_b = \eta_0 n_\gamma$.

$$\approx \frac{3}{4} \eta_0 X_n (1 - X_n) \left(\frac{m_d}{m_n m_p}\right)^{3/2} \frac{2}{\pi^2} (2\pi T)^{3/2} \zeta(3) \exp\left(\frac{B_d}{T}\right), \tag{3.141c}$$

Approximated $X_p + X_n \approx 1$, ignoring heavier nuclides.

Does approximating $X_p + X_n \approx 1$ not ignore deuterium as well? Or rather: by the definition given before $X_n + X_p = 1$ is exact, and if we do normalize by $n_n + n_p + n_d + \dots$ we should specify...

which describes the *deuterium bottleneck*: similarly to hydrogen recombination, the presence of many photons for each nuclide keeps compound particles from forming for quite a long time, and this precludes the formation of Helium.

We can approximate this as

$$X_d \approx X_n (1 - X_n) \exp\left(-29.33 + \frac{25.82}{T_9} - \frac{3}{2} \log T_9 + \log\left(\Omega_{0b} h^2\right)\right), \tag{3.142}$$

where $T_9 = T/10^9$ K. The factor $\Omega_{0b} h^2$ comes from the scaling of the baryon-to-photon ratio η (equation (3.119)).

Pacciani [Pac18b] says we use the exponential decay of the neutron population to find this formula: how does it come in, though?

Almost all the deuterium which is formed then turns into ^{4}He, so since those nuclei are as heavy as four neutrons, and two neutrons are needed to make each we can estimate that the helium mass fraction will be given by $Y \approx 2X_n \approx 0.25$. The calculation is quite rough, in order to do it properly we should use the Boltzmann equation.

Very rough calculation… the corrective factor of 0.8 is kind of thrown in there, also I think using $\exp(-1.5)$ instead of $\exp(-1.3)$ helps nudging the number and hides just how rough this is

This provides an experimental bound for many parameters: the lifetime of the neutron, for example, cannot be much shorter since otherwise too many would have decayed before forming deuterium.

The reaction rate Γ for weak interactions, as we have discussed earlier, is given by $\Gamma \sim G_F^2 T^5 \approx T^5/\tau_n$, where the Fermi coupling constant G_F is connected to the lifetime of the neutron τ_n since that is also a weak-interaction process.

Then, the moment of decoupling $\Gamma \sim H$ also constrains τ_n, since it happens at a temperature $T_{dv} \sim G_F^{-2/3} m_P^{-1/3} \approx \tau_n^{4/3} m_P^{-1/3}$.[26]

Increasing the lifetime of neutrons decreases the amount of He-4 in the universe.

We know that the Hubble parameter is given by

$$H^2 = \frac{8\pi G}{3} \underbrace{\frac{\pi^2}{30} g_*(T) T^4}_{\rho=\rho_r}, \tag{3.143}$$

so $H(T) \sim \sqrt{g_*(T)}$: this means that if we add more particles to our model which are coupled in the $\sim$ MeV range, thus increasing g_* at that temperature, we also increase H, which means that the moment at which $\Gamma < H$ comes about earlier, meaning that we get more Helium. This is very useful constraint on our models; for instance, any dark matter particle we hypothesize needs to not "break baryogenesis".

Decreasing the baryon fraction Ω_b inhibits the production of deuterium, since then there are more photons to disassociate them. We find that the model fits observation as long as $0.011h^{-2} \leq \Omega_{0b} \leq 0.25h^{-2}$; most people agree that we are near the upper bound of this range.

There is also another parameter, $m_P \sim G^{-1/2}$: modified gravity theories often predict variations of the gravitational constant with time, but since this appears in our calculations we can constrain its value at that stage in the early universe.

[26] One might think that the lifetime of the neutron would be a well-established result in Earthly particle physics, but in fact there is a conflict between two different kinds of measurements [Wol], so while the value is roughly known the error bars of these measurements do not overlap: this is the "neutron lifetime puzzle".

3.8 Dark Matter dynamics

We have described how dark matter came to be accepted as a necessary component of the matter content of the universe, now we wish to understand what are its basic properties.

Observationally, dark matter has no relevant electromagnetic interactions: in the low-redshift universe it only interacts gravitationally (it is *decoupled*), and is able to cluster.

This allows us to distinguish it from dark energy, which instead is uniformly distributed.

Models of dark matter can be classified into **Hot and Cold dark matter**: HDM and CDM. We can also have *Warm* dark matter, which has intermediate properties.

In **HDM**, particles decoupled while they were ultrarelativistic, so they have very high thermal motion. They move fast, and tend to destroy gravitational potential wells in which they might settle by moving out of them, and thus decreasing the quantity of matter there. Neutrinos were thought to be Dark Matter, and would have been classified as HDM.

They do this on scales comparable to the maximum distance travelled by them: this is calculated as vt, where v is their average thermal velocity, and t is the age of the universe.

This means that the structures formed in the presence of HDM are larger than $10^{15} M_\odot$; however this seems to conflict with observation, since we also see smaller structures! The hypothesis is then that they were formed later, by fragmentation: this is known as the *top-down approach*.

The top-down approach, however, is falsified by the observation of high-redshift quasars combined with the scale of the anisotropies of the CMB: in order to account fo high-redshift small-scale structures (we have seen stars at $z \sim 20$!) we would have to increase the amplitude of the anisotropies to a scale which is not compatible to the anisotropies we see in the CMB. This does not mean that HDM cannot exist, but it cannot make up most of the observed DM density.

CDM, on the other hand, is dark matter which decoupled at a time when it was already nonrelativistic. A *bottom-up* approach to structure formation is compatible with the existence of CDM. This is the kind of dark matter which appears in the currently-accepted ΛCDM model of cosmology.

Although it is the best model to date, there are also issues with CDM, which are current research topics. Dark matter is distributed in *halos*, gravitationally-bound regions which contain dark matter and are currently decoupled from cosmic expansion.

One issue is that simulations predict small-scale halos in the larger ones, which are however not observed. Also, the density profiles from simulations do not seem to match what is observed: they have a cusp in the center of the halo, while observations suggest that density profiles are flat in the center.

There are many proposed solutions to these problems: for example, dark matter could self-interact in ways other than gravitational, or it could be warm.

3.8.1 The Boltzmann equation and decoupling

In order to understand how Dark Matter evolves, let us introduce the main equation used to analyze non-equilibrium phenomena: the **Boltzmann equation**. It describes the evolution of the phase space distribution of particles, f. Its compact form is:

$$\mathbb{L}[f] = \mathbb{C}[f] \,, \tag{3.144}$$

where $\mathbb{L}$ is the *Liouville operator*, which is a total derivative of the phase space density with respect to changes in time, position as well as momentum; while $\mathbb{C}$ is the *collision operator*, which describes the effect (in terms of variation per unit time) on the phase space distribution of collisions between particles — it is written in terms of scattering matrices. In general, we should consider the phase space distribution of all the particle species, writing an equation for each of theses f_i; now we will treat a single particle species for simplicity.

This basic form is quite general: it can be used both for a simple Newtonian calculation and to include general-relativistic and quantum-mechanical effects.

We start with the Newtonian description: the phase space has position, momentum and time as coordinates, and a density function $f(\vec{q}, \vec{p}, t)$ is defined on it.

The classical form of the Liouville operator is the *convective derivative*,

$$\mathbb{L} = \frac{D}{Dt} = \frac{\partial}{\partial t} + \frac{\mathrm{d}\vec{x}}{\mathrm{d}t} \cdot \nabla_x + \frac{\mathrm{d}\vec{p}}{\mathrm{d}t} \cdot \nabla_p \tag{3.145a}$$

$$= \frac{\partial}{\partial t} + \vec{v} \cdot \nabla_x + \frac{\vec{F}}{m} \cdot \nabla_p \,. \tag{3.145b}$$

This equation describes motion in much more detail than, say, the Navier-Stokes equations, since it allows for arbitrary momentum *at each point* in space. We can recover the N-S equations, as well as the energy and mass conservation equations, by taking *moments* of the Boltzmann equation: we multiply it by $m\vec{v}$, or $mv^2/2$, or m, and integrate in $\mathrm{d}^3 p$.

In the Newtonian case, the force term $\vec{F}$ includes gravity, which in GR instead is "geometrized", meaning that it is included in the inertial motion of the particles. The relativistic version of the Liouville operator is written in terms of position x^α and momentum $p^\alpha = \mathrm{d}x^\alpha / \mathrm{d}\lambda$

$$\mathbb{L} = \frac{\mathrm{d}x^\alpha}{\mathrm{d}\lambda} \frac{\partial}{\partial x^\alpha} + \frac{\mathrm{d}p^\alpha}{\mathrm{d}\lambda} \frac{\partial}{\partial p^\alpha} \tag{3.146}$$

$$= p^\alpha \partial_\alpha - \Gamma^\alpha_{\beta\gamma} p^\beta p^\gamma \frac{\partial}{\partial p^\alpha} \,. \tag{3.147}$$

Is the momentum supposed to be dimensionless or not? Setting $\mathrm{d}x^\alpha / \mathrm{d}\lambda = p^\alpha$ as well as $p^2 = m^2$ seems contradictory…

This is because, since any particle must move along geodesics, p^α must satisfy the geodesic equation:

$$\frac{Dp^\alpha}{Dt} = \frac{\mathrm{d}p^\alpha}{\mathrm{d}\lambda} + \Gamma^\alpha_{\beta\gamma} p^\beta p^\gamma = 0 \,. \tag{3.148}$$

We require that the particles are *on shell*:

$$g^{\alpha\beta} p_\alpha p_\beta = m^2 \,, \tag{3.149}$$

meaning that we are neglecting the uncertainty principle and our discussion is relativistic but classical (not fully quantum-mechanical).

Let us make this explicit, using the Christoffel symbols of the FLRW metric and assuming isotropy and homogeneity (which together imply that the phase space density f only depends on t and $p^0 = E$): many terms vanish, and we are left with

$$\mathbb{L}[f] = p^0 \partial_t f - \Gamma^0_{ij} p^i p^j \frac{\partial f}{\partial E} = E \partial_t f - \frac{\dot{a}}{a} |\vec{p}|^2 \frac{\partial f}{\partial E}, \tag{3.150}$$

since the relevant Christoffel symbols (in Cartesian coordinates) are

$$\Gamma^0_{ij} = \frac{\dot{a}}{a} \delta_{ij}. \tag{3.151}$$

So, we can write the Boltzmann equation as

$$\frac{\partial f}{\partial t} - \frac{\dot{a}}{a} \frac{p^2}{E} \frac{\partial f}{\partial E} = \frac{\mathbb{C}[f]}{E}. \tag{3.152}$$

Let us now do what is also done in the classical case and integrate in $\mathrm{d}^3\vec{p}$: we will lose detail in the description but gain in computability. First, recall the definition of the number density from the phase space distribution:

$$n(t) = \frac{g}{(2\pi)^3} \int \mathrm{d}^3\vec{p}\, f(|\vec{p}|, t), \tag{3.153}$$

which appears in the first term if we integrate the Boltzmann equation (also multiplying by $g/(2\pi)^3$):

$$\frac{\partial n}{\partial t} - \frac{g}{(2\pi)^3} \frac{\dot{a}}{a} \int \mathrm{d}^3\vec{p}\, \frac{p^2}{E} \frac{\mathrm{d}f}{\mathrm{d}E} = \frac{g}{(2\pi)^3} \int \mathrm{d}^3 p\, \frac{1}{E} \mathbb{C}[f]. \tag{3.154}$$

The number density actually also appears in the second term: we can manipulate it as

$$\begin{aligned}
\int \mathrm{d}^3 p\, \frac{p^2}{E} \frac{\partial f}{\partial E} &= 2 \int \mathrm{d}^3 p\, p^2 \frac{\partial f}{\partial (E^2)} = 2 \int \mathrm{d}^3 p\, p^2 \frac{\partial f}{\partial (p^2)} && \text{(3.155a)} \\
&= \int \mathrm{d}^3 p\, p \frac{\partial f}{\partial p} = 4\pi \int_0^\infty \mathrm{d}p\, p^3 \frac{\partial f}{\partial p} && \text{(3.155b)} \\
&= 4\pi p^3 f \Big|_0^\infty - 4\pi \int_0^\infty \mathrm{d}p\, 3p^2 f = -3 \int \mathrm{d}^3 p\, f = -3n \frac{(2\pi)^3}{g}. && \text{(3.155c)}
\end{aligned}$$

$\mathrm{d}(E^2) = 2E\,\mathrm{d}E$ and similarly for $\mathrm{d}(p^2)$.

The boundary term vanishes since at 0 we have $p = 0$, and at (momentum) infinity we have $f = 0$ (since otherwise the energy would diverge). We are also moving back and forth between integrals in $\mathrm{d}^3 p$ and in $\mathrm{d}p$ times 4π, which we can do because of isotropy.

So, we can see that the left-hand side is equal to $\dot{n} + 3\dot{a}n/a$, and we have the **cosmological Boltzmann equation**

$$\dot{n} + 3\frac{\dot{a}}{a} n = \frac{g}{(2\pi)^3} \int \mathrm{d}^3 p\, \frac{1}{E} \hat{\mathbb{C}}[f]. \tag{3.156}$$

Right away we can see that this makes sense in the decoupling limit: if there are no collisions the right-hand side must vanish, so we are left with

$$\dot{n} + 3\frac{\dot{a}}{a}n = \frac{\mathrm{d}}{\mathrm{d}t}\left(na^3\right) = 0\,, \tag{3.157}$$

meaning that $n \propto a^{-3}$: this is the usual scaling of the number density, so our approach is working.

Now we need to understand what the right-hand side looks if particle species are coupled. This is difficult in general, we will give an expression using "bulk" parameters:

$$\frac{g}{(2\pi)^3}\int \frac{\mathrm{d}^3p}{E}\hat{\mathbb{C}}[f] = \Psi - \Gamma_A n = \Psi - \langle\sigma v\rangle\, n^2\,, \tag{3.158}$$

where Ψ is the rate of creation of particles per unit volume, while Γ_A is the rate of annihilation. Note that Ψ has the dimensions $\mathrm{s}^{-1}\mathrm{m}^{-3}$, while Γ_A has the dimensions s^{-1}, so it needs to be multiplied by n in order to be dimensionally consistent. Physically speaking, the reason the term has that form is that the rate of annihilation must be proportional to the number density of particles there are at the moment.

The annihilation rate is found in terms of the cross section of the process, which we must average over all the momentum distribution of the particles: this is the reason for the appearance of $\langle\sigma v\rangle$.

At equilibrium, the left-hand side is equal to zero. The number density will be given by some equilibrium value, n_{eq}: inserting this in the right-hand side, which must also vanish, we find $\Psi = \langle\sigma v\rangle\, n_{\text{eq}}^2$. This result can then be used in general: the right-hand side is written as $\langle\sigma v\rangle\left(n_{\text{eq}}^2 - n^2\right)$.

This means that the equation reads

$$\dot{n} + 3\frac{\dot{a}}{a}n = \langle\sigma_A v\rangle\left(n_{\text{eq}}^2 - n^2\right). \tag{3.159}$$

Since the number density n itself is not conserved, we define the *comoving* number density, which is conserved if there is equilibrium

$$n_C = n\left(\frac{a}{a_0}\right)^3, \tag{3.160}$$

for some arbitrary initial scale factor a_0. With this, we can simplify the left-hand side:

$$\dot{n} + 3\frac{\dot{a}}{a}n = \frac{\mathrm{d}}{\mathrm{d}t}\left[n_C\frac{a_0^3}{a^3}\right] + 3\frac{\dot{a}}{a}n_C\frac{a_0^3}{a^3} = \dot{n}_C\frac{a_0^3}{a^3} + n_C\left(-\frac{3\dot{a}}{a^2}\right)\frac{a_0^3}{a^2} + 3\frac{\dot{a}}{a}n_C\frac{a_0^3}{a^3} \tag{3.161}$$

$$= \dot{n}_C\left(\frac{a_0}{a}\right)^3. \tag{3.162}$$

Similarly, we can express the right-hand side in terms of comoving densities: $\langle\sigma_A v\rangle$ is the same, while

$$n_{\text{eq}}^2 - n^2 = \left(n_{\text{C,eq}}^2 - n_{\text{C}}^2\right)\frac{a_0^6}{a^6} = n_{\text{C,eq}}^2\frac{a_0^6}{a^6}\left[1 - \frac{n_{\text{C}}^2}{n_{\text{C,eq}}^2}\right], \tag{3.163}$$

so the equation will read:

$$\dot{n}_C \frac{a_0^3}{a^3} = -\langle \sigma_A v \rangle\, n_{C,\text{eq}}^2 \frac{a_0^6}{a^6} \left[\underbrace{\frac{n_C^2}{n_{C,\text{eq}}^2}}_{\text{The } a^3 \text{ factors simplify}} -1 \right] \tag{3.164}$$

$$\dot{n}_C = -\underbrace{\langle \sigma_A v \rangle \left(\frac{a_0}{a} \right)^3 n_{C,\text{eq}}^2}_{=n_{C,\text{eq}} n_{\text{eq}}} \left[\frac{n^2}{n_{\text{eq}}^2} - 1 \right]. \tag{3.165}$$

We want to write this in a more intuitive way; the derivative with respect to time can be expressed in terms of the scale factor, as

$$\frac{\mathrm{d}}{\mathrm{d}t} = \dot{a} \frac{\mathrm{d}}{\mathrm{d}a} = Ha \frac{\mathrm{d}}{\mathrm{d}a}\,. \tag{3.166}$$

With it, we can express the equation as

$$\frac{a}{n_{C,\,\text{eq}}} \frac{\mathrm{d}n_C}{\mathrm{d}a} = -\frac{\langle \sigma_A v \rangle\, n_{\text{eq}}}{H} \left[\left(\frac{n}{n_{\text{eq}}} \right)^2 - 1 \right]. \tag{3.167}$$

The ratio before the parenthesis has an intuitive physical meaning: the characteristic time of the collisions which can annihilate this particle species is $\tau_{\text{coll}} = 1/\Gamma = 1/(\langle \sigma_A v \rangle\, n_{\text{eq}})$, while the timescale of the expansion of the universe is $\tau_{\text{exp}} = 1/H$, we get

$$\boxed{\frac{a}{n_{C,\,\text{eq}}} \frac{\mathrm{d}n_C}{\mathrm{d}a} = -\frac{\tau_{\text{exp}}}{\tau_{\text{coll}}} \left[\left(\frac{n}{n_{\text{eq}}} \right)^2 - 1 \right].} \tag{3.168}$$

This allows us to characterize decoupling in a much more specific way than before.

1. If $\Gamma \gg H$, then $\tau_{\text{exp}} \gg \tau_{\text{coll}}$, therefore $n \approx n_{\text{eq}}$, which also implies $n_C = n_{C,\,\text{eq}}$ (this quantity can vary!). This is the equilibrium case, the particles are **coupled**.

2. If $\Gamma \ll H$, then $\tau_{\text{exp}} \ll \tau_{\text{coll}}$ we have **decoupling**, and $n_C = \text{const}$, therefore $n \propto a^{-3}$.

Pacciani [Pac18b] writes that in the coupled case the comoving density is constant, but this will not be the case in general: allowed annihilation processes may change over time, as other particle species decouple!

This approach is much more powerful than the qualitative one we gave earlier, since while the limiting cases are the same this equation allows us to also treat all the intermediate situations.

Let us apply this to both HDM and CDM.

3.8.2 HDM density estimate

Neutrinos are a candidate for HDM: we know that at temperatures below $T_d = 1\,\text{MeV}$ they decoupled, after which — as we have seen earlier — their temperature[27] evolved like:

$$T_\nu = \left(\frac{g_{*\text{after decoupling}}}{g_{*\text{before decoupling}}} \right)^{1/3} T_\gamma \,, \tag{3.169}$$

and so this scaled like $T_\nu \propto a^{-1}$, while their number density scaled like $n_\nu \propto a^{-3} \propto T_\nu^3$.

There is nothing special about neutrinos: we can apply the same line of reasoning to a generic HDM species x, whose number density today will be then given by

$$n_{0x} = B g_x \frac{\zeta(3)}{\pi^2} T_{0x}^3 \,, \tag{3.170}$$

where the factor B accounts for the statistics: it is 1 for bosons, $3/4$ for fermions. The parameter g_x is the number of degrees of freedom *of the particle species* x,

We can rescale this in terms of the photon number density, which is given by the same expression, with $B = 1$ and $g_x = 2$: using the

$$\frac{n_{0x}}{n_{0\gamma}} = \frac{B g_x \frac{\zeta(3)}{\pi^2} T_{0x}^3}{2 \frac{\zeta(3)}{\pi^2} T_{0\gamma}^3} \tag{3.171}$$

$$n_{0x} = \frac{B}{2} n_{0\gamma} g_x \frac{g_{*0}}{g_{*dx}} \,, \tag{3.172}$$

where g_{*0} is the current amount of effective degrees of freedom, while g_{*dx} is the same quantity, computed at the decoupling time of particle x.

It should not be g_{*0} though, right? we compute g_* on the two boundaries of the transition, the effective dof *right now* are irrelevant, I'd think.

The energy density, as long as today the particles in question are nonrelativistic,[28] is given by $\rho_{0x} = m_x n_{0x}$, so we can write

$$\rho_{0x} = \frac{B}{2} m_x n_{0\gamma} g_x \frac{g_{*0x}}{g_{*dx}} \,, \tag{3.173}$$

[27] We call it "temperature" for clarity, but properly speaking it is not one, since ever since decoupling neutrinos are not thermal. It is better understood as the "temperature parameter" of the neutrinos' phase space distribution.

[28] Which we assume they are. This is not meant to be interpreted as an experimental fact — for neutrinos, say, we do not actually know what their mass is, so we are not sure, although given their mass differences (which can be inferred from neutrino oscillations) at the current temperature $T_{0\nu} \sim 200\,\mu\text{eV}$ at least some neutrino species must be nonrelativistic.

Rather, the point is that if the mass of one of these HDM particles were so low that they were relativistic today than their energy density would be very low, comparable to that of photons, which as we have seen earlier is negligible: they could not constitute the large fraction of the critical density ρ_C that we know DM constitutes. In order for HDM to have a chance to be a significant constitutent of DM it must be nonrelativistic today.

therefore the mass fraction of the HDM particle x today will be roughly

$$\Omega_{0x} h^2 \approx \frac{\rho_{0x}}{\rho_{0c}} h^2 \approx 2B g_x \frac{g_{*0}}{g_{*dx}} \frac{m_x}{10^2\,\text{eV}} \,. \tag{3.174}$$

This, together with what we know the Ω of dark matter to be, allows us to check whether a candidate for HDM is viable or not, based on its mass and on when it decouples. We can already see that if the mass is lower than a few eV the particle cannot make up most of the DM budget.

3.8.3 CDM density estimate

CDM is made of particles which were already nonrelativistic when they decoupled, so we can describe them using Boltzmann statistics: we keep referring to the DM candidate as x, so at the temperature of decoupling T_{dx} we have

$$n_x(T_{dx}) = g_x \left(\frac{m_x T_{dx}}{2\pi} \right)^{3/2} \exp\left(-\frac{m_x}{T_{dx}} \right) , \tag{3.175}$$

after which the density will scale like a^{-3}, so[29]

$$n_{0x} = n_x(T_{dx}) \left(\frac{a(T_{dx})}{a_0} \right)^3 = n(T_{dx}) \frac{g_{*0}}{g_{*x}} \left(\frac{T_{0\gamma}}{T_{dx}} \right)^3 . \tag{3.176}$$

The difficulty lies in determining the decoupling temperature T_{dx}, which is when the collision and expansion timescales are equal.

The first Friedmann equation combined with the expression for the energy density (of radiation, but corrected according to the effective degrees of freedom at that time) tells us

$$H^2(T_{dx}) = \frac{8\pi G}{3} g_{*dx} \frac{\pi^2}{30} T_{dx}^4 , \tag{3.177}$$

which we can use to estimate the expansion timescale $\tau_{\text{exp}} = 1/H$:

$$\tau_{\text{exp}} \approx 0.6 g_{*dx}^{-1/2} \frac{m_{\text{pl}}}{T_{dx}^2} \,. \tag{3.178}$$

$0.6 \approx \left(\frac{8\pi}{3} \frac{\pi^2}{30} \right)^{-1/2}$.

Pacciani [Pac18b] writes 0.3, am I getting the calculation wrong?

Now, let us estimate the collision timescale: we know that its inverse is $\Gamma = n \langle \sigma_A v \rangle$, and it is a fact from particle physics that the average cross section scales with the temperature like:

$$\langle \sigma_A v \rangle = \sigma_0 \left(\frac{T}{m_x} \right)^N , \tag{3.179}$$

[29] Applying the "updated" version of Tolman's law, $T a g_{*s}^{1/3} = \text{const}$.

where $N = 0$ or 1, while σ_0 is a constant characteristic cross section of the process. So, the collision timescale is

$$\tau_{\text{coll}}(T_{dx}) = \left(n(T_{dx})\sigma_0 \left(\frac{T_{dx}}{m_x} \right)^N \right)^{-1} . \tag{3.180}$$

Equating the two timescales we get the following equation:

$$\left(n(T_{dx})\sigma_0 \left(\frac{T_{dx}}{m_x} \right)^N \right)^{-1} = 0.6 g_*^{-1/2} \frac{m_{\text{pl}}}{T_{dx}^2} , \tag{3.181}$$

which is transcendental in T, since we have an exponential as well as a polynomial term in the expression for $n_x(T_{dx})$.

We can solve it iteratively, in terms of the parameter $x_{dx} = m_x / T_{dx}$, which are assuming to be much larger than one (in order for the procedure we have done so far to be valid, and in order for x to be CDM): this allows us to select the physical solution to the equation among the nonphysical ones.

The solution, after the second iteration, is found to be something like:

$$x_{dx} = \log \left(0.038 \frac{g_x}{g_{*xd}^{1/2}} m_{\text{pl}} m_x \sigma_0 \right) - \left(N - \frac{1}{2} \right) \log \log \left(0.038 \frac{g_x}{g_{*xd}^{1/2}} m_{\text{pl}} m_x \sigma_0 \right) . \tag{3.182}$$

From this we can determine the contribution of this CDM particle to the current energy density.

Qualitatively, we find that there is a very significant dependence on the mass of the particle (since we have an exponential suppression in its density) and on the way it interacts (σ_0).

Chapter 4

Stellar Astrophysics

4.1 Stellar formation

We will now discuss the formation of stars, which started happening at a redshift $z \sim 20$. We shall do so in a Newtonian approximation, neglecting the expansion of the universe — later we will discuss how cosmology affects this.

A star is a gravitationally-bound sphere of plasma, inside which fusion occurs. Stars form from the gravitational collapse of instabilities, which is followed by an increase in temperature and pressure from the release of gravitational energy.

We will model this through simple assumptions, since they already give a good picture of what this collapse looks like. Two forces are at play: gravity and pressure.[1] As we will see, the gravitational force initially dominates and compresses the material up to a certain point, at which the pressure prevents it from going further.

We start with a spherically symmetric region of baryonic matter, characterized by a density $\rho(r)$: the mass enclosed in a radius r is

$$m(r) = \int_0^r 4\pi \widetilde{r}^2 \rho(\widetilde{r}) \, \mathrm{d}\widetilde{r} \,. \tag{4.1}$$

The modulus of the gravitational acceleration of the material at a radial coordinate r can be calculated from Gauss' theorem:

$$g(r) = \frac{Gm(r)}{r^2} \,. \tag{4.2}$$

Let us now consider a spherical shell at a radius r, with its enclosed mass $\Delta M = \rho(r)\Delta A \Delta r$, where $\Delta A = 4\pi r^2$. Let us denote P as the pressure at the inner surface, and $P + \Delta P$ the pressure at the outer surface. Then, the net force on the surface is given by

$$(P + \Delta P)\Delta A - P\Delta A = \left(P(r) + \frac{\mathrm{d}P}{\mathrm{d}r} \Delta r \right) \Delta A - P(r)\Delta A = \frac{\mathrm{d}P}{\mathrm{d}r} \Delta A \Delta r = \frac{\mathrm{d}P}{\mathrm{d}r} \frac{\Delta M}{\rho(r)} \,. \tag{4.3}$$

[1] In this section we will neglect the Pauli exclusion principle, which does not allow fermionic matter to compress beyond a certain point. We will come back to this point when we discuss the Chandrasekhar mass.

Note that this is an inward force if ΔP is positive (and so $\mathrm{d}P/\mathrm{d}r$ also is), since then there is more pressure outside than inside.

The equation of motion of the spherical shell is given by $ma = F$:

$$-\Delta M \ddot{r} = \Delta M g(r) + \frac{\mathrm{d}P}{\mathrm{d}r} \frac{\Delta M}{\rho(r)} \tag{4.4}$$

$$-\ddot{r} = \frac{Gm(r)}{r^2} + \frac{1}{\rho(r)} \frac{\mathrm{d}P}{\mathrm{d}r}, \tag{4.5}$$

where the minus sign comes from the two forces are positive if they are directed inward. This means that, in order to achieve equilibrium ($\ddot{r} = 0$), the pressure gradient $\mathrm{d}P/\mathrm{d}r$ must be negative, since $m(r)$ can never be.

Now we shall give two estimates about stellar formation: the first is the **free-fall timescale**, in which we ignore pressure forces in order to ballpark the time taken for the matter to fall onto itself.

Then, we will study the equilibrium configuration of the star, in order to understand what are the conditions under which it is actually gravitationally bound (that is, with negative total energy).

4.1.1 The freefall timescale

We are ignoring pressure (or, equivalently, assuming that it is constant), so the equation of motion reads

$$-\ddot{r} = g(r). \tag{4.6}$$

This will not generally be the case, but let us suppose that the collapse is *orderly*: the ordering of the layers stays the same as they fall.

We use the energy integral instead of directly solving the differential equation, that is, we impose energy conservation. Computing the total energy (kinetic plus potential) at the initial radius of the cloud, r_0 (at which the gas is stationary), and at a radius r we get

$$-\frac{Gm_0}{r_0} = \frac{1}{2}\left(\frac{\mathrm{d}r}{\mathrm{d}t}\right)^2 - \frac{Gm_0}{r} \tag{4.7}$$

$$\frac{1}{2}\left(\frac{\mathrm{d}r}{\mathrm{d}t}\right)^2 = Gm_0\left(\frac{1}{r} - \frac{1}{r_0}\right). \tag{4.8}$$

This way, we have found a first-order ODE instead of a second-order one. Note that the potential energy at a radius r is computed using m_0 since as the layer falls the other layers below it are still below it, so the mass inside the layer is always equal to the initial one.

We can now directly compute the freefall time by integrating from r_0 to r:

$$t_{\text{free fall}} = \int_{r_0}^{0} \mathrm{d}r \, \frac{\mathrm{d}t}{\mathrm{d}r} = -\int_{r_0}^{0} \mathrm{d}r \, \frac{1}{\sqrt{2Gm_0}} \left(\frac{1}{r} - \frac{1}{r_0}\right)^{-1/2}, \tag{4.9}$$

where we have a minus sign since, when simplifying the square of the derivative $\dot{r}^2$ we must choose the negative sign: $\mathrm{d}r/\mathrm{d}t < 0$, since the material is *infalling*.

We can then change variables to $x = r/r_0$ (with $\mathrm{d}r = r_0\,\mathrm{d}x$) and switch the bounds of integration, recovering the positive sign:

$$t_{\text{free fall}} = \frac{1}{\sqrt{2Gm_0}} \int_0^1 r_0\,\mathrm{d}x \left[\frac{1}{r_0}\left(\frac{1}{x} - 1\right)\right]^{-1/2} \tag{4.10}$$

$$= \sqrt{\frac{r_0^3}{2Gm_0}} \int_0^1 \mathrm{d}x \sqrt{\frac{x}{1-x}} = \frac{\pi}{2}\sqrt{\frac{r_0^3}{2Gm_0}}\,. \tag{4.11}$$

Although the integrand diverges for $x \to 1$ (meaning, at large r) the integral converges to $\pi/2$.[2]

The average density is given by $\bar{\rho} = m_0/\left(4\pi r_0^3/3\right)$, which we can insert into our expression to get

$$t_{\text{free fall}} = \sqrt{\frac{3\pi}{32G\bar{\rho}}} \approx 0.54 G^{-1/2}\bar{\rho}^{-1/2}\,. \tag{4.13}$$

Comparison with the expansion timescale We might be tempted to ignore the expansion of the universe in these calculations: we know that

$$H^2 = \frac{8\pi G}{3}\bar{\rho}\,, \tag{4.14}$$

where we wrote $\bar{\rho}$ to mean that this holds on the scales at which homogeneity applies; if we consider smaller scales we must take a spatial average of the density. In the Newtonian (matter-dominated and flat) case we know that $a(t) \propto t^{2/3}$ and $H = 2/(3t)$.

So, we can substitute this expression to find the expansion timescale

$$\frac{4}{3t^2} = \frac{8\pi G}{3}\bar{\rho} \implies t^2 = \frac{4}{9}\frac{3}{8\pi G}\bar{\rho}^{-1}\,, \tag{4.15}$$

so

$$t_{\exp} \approx 0.23 G^{-1/2}\bar{\rho}^{-1/2}\,. \tag{4.16}$$

If the universe was perfectly homogeneous we would then expect structure formation to be forbidden: however, if there are some over-dense regions to start with, their characteristic freefall time can become lower than the chracteristic expansion time of the universe.

But still, shouldn't we consider expansion in the computation? By how much are we getting it wrong?

[2] The integral can be computed with the substitution $x = \sin^2\theta$ and then $y = \cos\theta$:

$$\int_0^1 \sqrt{\frac{x}{1-x}}\,\mathrm{d}x = \int_0^{\pi/2} \sqrt{\frac{\sin^2\theta}{\cos^2\theta}} 2\sin\theta\cos\theta\,\mathrm{d}\theta = \int_0^{\pi/2} \sin^2\theta\,\mathrm{d}\theta = \int_0^1 \sqrt{1-y^2}\,\mathrm{d}y = \frac{\pi}{2}\,. \tag{4.12}$$

4.1.2 Hydrostatic equilibrium

At equilibrium the stellar layers are static: $\ddot{r} = 0$, so the equation of motion reads

$$-\ddot{r} = 0 = g(r) + \frac{1}{\rho(r)}\frac{\mathrm{d}P}{\mathrm{d}r} \implies \frac{\mathrm{d}P}{\mathrm{d}r} = -G\frac{m(r)\rho(r)}{r^2}\,. \tag{4.17}$$

We multiply both sides by $4\pi r^3$ and integrate in $\mathrm{d}r$ from the core ($r = 0$) to the surface of the star, $r = R$:

$$\int_0^R \mathrm{d}r\, 4\pi r^3 \frac{\mathrm{d}P}{\mathrm{d}r} = -G\int_0^R \frac{m(r)\rho(r)4\pi r^2}{r}\,\mathrm{d}r\,, \tag{4.18}$$

and we can change variables: $\rho(r)4\pi r^2\,\mathrm{d}r = \mathrm{d}m$ (this is physically meaningful: it is the differential mass of the layer at a radius r), so we can identify the left-hand side with the total gravitational energy:

$$E_{\text{grav}} = -G\int \frac{m(r)\,\mathrm{d}m}{r}\,. \tag{4.19}$$

On the RHS, instead, we can integrate by parts:

$$\int_0^R \mathrm{d}r\, 4\pi r^3 \frac{\mathrm{d}P}{\mathrm{d}r} = \Big[P(r)4\pi r^3\Big]_0^R - 3\int_0^R \mathrm{d}r\, 4\pi r^2 P(r)\,, \tag{4.20}$$

where the boundary term vanishes: at the origin $r = 0$, at the surface (by definition of surface) $P = 0$.

We can better understand what this means if we divide and multiply by the volume $V(R) \equiv 4\pi R^3/3$:

$$-3\int_0^R \underbrace{\mathrm{d}r\, 4\pi r^2}_{\mathrm{d}V} P(r) = -3V(R)\underbrace{\int_0^R \frac{\mathrm{d}r\, 4\pi r^2 P(r)}{V(R)}}_{\langle P\rangle} = -3V(R)\langle P\rangle\,, \tag{4.21}$$

where we interpret the integral as a weighted average, so we get

$$E_{\text{grav}} = -3\langle P\rangle V \qquad \text{or} \qquad \langle P\rangle = -\frac{1}{3}\frac{E_{\text{grav}}}{V} = -\frac{1}{3}\rho_{GR}\,. \tag{4.22}$$

This is the **Virial Theorem**.

Now, the question we want to as is: is this equilibrium configuration **stable**? This is equivalent to asking whether the system is gravitationally bound, $E_{\text{grav}} < 0$, which as we have shown is equivalent to $\langle P\rangle > 0$.

In order to answer this question we shall use a statistical-mechanics, microscopic approach.

We consider a cubic box of volume $V = L^3$ with N particles inside it, each of which has a velocity $\vec{v} = (v_x, v_y, v_z)^\top$ and a momentum p. Let us select a face of the box, which we assume to be perpendicular to the x axis. Each particle will hit it with a frequency

$t^{-1} = v_x/2L$, and each time it does so it imparts upon it a momentum $2p_x$, since it is reflected backwards.

Summing over all the particles, the rate of momentum transfer (so, the force) in the direction x is given by

$$\frac{N}{2L} \left\langle 2p_x v_x \right\rangle \,, \tag{4.23}$$

so the pressure upon that face will be the force divided by the area of the face

$$P_x = \frac{N}{L} \left\langle p_x v_x \right\rangle \frac{1}{L^2} = \underbrace{\frac{N}{V}}_{=n} \left\langle p_x v_x \right\rangle \,. \tag{4.24}$$

This will be the same for each direction by isotropicity: $P_x = P_y = P_z$, and by the same argument we can write $\left\langle p_x v_x \right\rangle = \left\langle \vec{p} \cdot \vec{v} \right\rangle /3$: so

$$P = \frac{n}{3} \left\langle \vec{p} \cdot \vec{v} \right\rangle \,, \tag{4.25}$$

which, although we will not show it, generalizes to a configuration of any shape, and does not change if we consider quantum-mechanical or relativistic effects. This is a simple expression for the **equipartition theorem**, a crucial result in Hamiltonian mechanics.

Let us consider two limits: nonrelativistic and fully relativistic particles.

Nonrelativistic particles In this case, since $\gamma \approx 1$ the four-momentum of the particles is approximately

$$p^\mu = \begin{bmatrix} \gamma m c^2 \\ \gamma m \vec{v} \end{bmatrix} \approx \begin{bmatrix} m c^2 \\ m \vec{v} \end{bmatrix} \,, \tag{4.26}$$

so $\vec{p} = m\vec{v}$, which means $\left\langle \vec{p} \cdot \vec{v} \right\rangle = \left\langle m v^2 \right\rangle$.

Then, for a gas of nonrelativistic particles we can write the pressure as

$$P = \frac{n}{3} \left\langle m v^2 \right\rangle = \frac{2}{3} \rho_{E_K} \,, \tag{4.27}$$

where $\rho_{E_K} = nm \left\langle v^2/2 \right\rangle$ is the density of translational kinetic energy.

Combining this result with the fact that, as we have seen before, $\langle P \rangle = -\rho_{\text{grav}}/3$, we find

$$-\frac{1}{3} \rho_{\text{grav}} = \frac{2}{3} \rho_{E_K} \implies 2E_{\text{K}} + E_{\text{grav}} = 0 \,, \tag{4.28}$$

which is an alternate statement of the **nonrelativistic** case of the **virial theorem**.

The total energy is then given by $E_{\text{tot}} = E_k + E_{\text{grav}} = -E_k$: this means that in general the system will be **bound** — the kinetic energy is quadratic, so always positive — and that the hotter it is, the more bound it is.

Relativistic case In this case $v \approx c$, so $\langle p \cdot \vec{v} \rangle \approx pc$. We can apply the reasoning from before, but the density of translational kinetic energy is given by

$$\rho_{E_K} = n(E - mc^2) = n(\gamma - 1)mc^2 \approx \gamma mc^2 = npc\,, \tag{4.29}$$

so we have

$$P = \frac{n}{3} \langle \vec{p} \cdot \vec{v} \rangle = \frac{\rho_{E_k}}{3}\,. \tag{4.30}$$

Then, we can apply the same reasoning as the nonrelativistic case, with the difference of the missing factor 2: we then get

$$\rho_{E_k} + \rho_{\text{grav}} = 0 \implies E_{\text{grav}} + E_k = E_{\text{tot}} = 0\,, \tag{4.31}$$

so the system is **unbound**, it does not have any constraint preventing it from dissociating.

Adiabatic gas We have seen the limiting cases, now let us consider a slightly more general one: a gas undergoing an adiabatic transformation, such that PV^γ (with some real number γ) is constant.[3] We will show that this is equivalent to the equations of state considered in cosmology, where $P = w\rho$. This will allow us to characterize the gravitational stability of the to-be star depending on the equation of state of the gas.

We start by differentiating: $\mathrm{d}(PV^\gamma) = 0$, which means that we also have $\mathrm{d}(\log(PV^\gamma)) = 0$, which we can expand into

$$\mathrm{d}\log(V^\gamma) + \mathrm{d}\log(P) = \gamma \frac{\mathrm{d}V}{V} + \frac{\mathrm{d}P}{P} = 0\,, \tag{4.32}$$

so

$$-(\gamma - 1)P\,\mathrm{d}V = P\,\mathrm{d}V + V\,\mathrm{d}P = \mathrm{d}(PV)\;. \tag{4.33}$$

In an adiabatic transformation the entropy must not change: so, we can write

$$T\,\mathrm{d}S = \mathrm{d}E_{\text{in}} + P\,\mathrm{d}V = 0\,, \tag{4.34}$$

which we can then write using the relation we derived previously:

$$\mathrm{d}E_{\text{in}} = \frac{1}{\gamma - 1}\,\mathrm{d}(PV) \tag{4.35}$$

$$E_{\text{in}} = \frac{PV}{\gamma - 1} \tag{4.36}$$

$$P = (\gamma - 1)\frac{E_{\text{in}}}{V} = (\gamma - 1)\rho_{\text{in}}\,. \tag{4.37}$$

We can then see that if we impose that the transformation be adiabatic, we find the equation of state $P = w\rho$, with $\gamma - 1 = w$.

[3] A more realistic model would allow γ to vary, which it definitely does in the stages of stellar formation and evolution and even across a single transformation. We will not, however, get that deep in the weeds.

Using the fact that, as we have shown before, $P = -\rho_{\text{grav}}/3$, this means

$$-\frac{\rho_{\text{grav}}}{3} = (\gamma - 1)\rho_{\text{in}} \implies 3(\gamma - 1)E_{\text{in}} + E_{\text{gr}} = 0\,, \tag{4.38}$$

which, together with the fact that the total energy of the star after the collapse is the initial energy plus the (negative) gravitational binding energy: $E_{\text{tot}} = E_{\text{in}} + E_{\text{gr}}$, so

$$E_{\text{tot}} = -(3\gamma - 4)E_{\text{in}}\,, \tag{4.39}$$

which means that $\gamma > 4/3$ characterizes a bound system, while $\gamma < 4/3$ characterizes a free system. This is consistent with what we have seen before: the limiting case $\gamma = 4/3$ is equivalent to $w = 1/3$, the equation of state of radiation (or ultrarelativistic matter), which as we have already seen is unbound.

From classical thermodynamics we know that, for instance, a monoatomic gas has $\gamma = 5/3$.

4.2 Jeans instability

Let us now try to understand the conditions under which a cloud of gas may become unstable and collapse onto itself to form a star (or a planet, for that matter).

In general, the gravitational potential energy of a body whose characteristic size is R and whose mass is M is given by

$$E_{\text{grav}} = -\int_{x,y\in V} \mathrm{d}^3x\, \mathrm{d}^3y\, \rho(x)\rho(y)\frac{G}{|x-y|} = -f\frac{GM^2}{R}\,, \tag{4.40}$$

where f is a numerical factor depending on the mass distribution. If the object at hand is uniform-density sphere, we have $f = 3/5$. In general, the factor is of order 1.

The kinetic component of the energy, on the other hand, is

$$E_{\text{K}} = \frac{3}{2}Nk_BT\,. \tag{4.41}$$

The gravitational cloud is unstable the gravitational energy is larger than the kinetic energy:

Why should this be? The way Keeton [Kee14] discusses it makes more sense to me: he studies the response of the total energy to a decrease in radius, and checks that it is positive; it is not the same as what we are doing here!

$$f\frac{GM^2}{R} > \frac{3}{2}Nk_BT\,, \tag{4.42}$$

and the Jeans mass, M_J, corresponds to the boundary of the stability region: the number of particles, N, depends on it as $N = M_J/\overline{m}$, where $\overline{m}$ is the average particle mass.

The criterion then reads:

$$f\frac{gM_J^2}{R} = \frac{3}{2}\frac{M_J}{\bar{m}}k_B T \tag{4.43}$$

$$M_J = \frac{3}{2}\frac{k_B T}{G\bar{m}}R\,, \tag{4.44}$$

where we set $f = 1$, since we are only interested in an order-of-magnitude calculation.

As usual, we want to reframe our result in terms of densities: the Jeans mass corresponds to a Jeans density times the volume of the sphere:

$$M_J = \frac{4\pi}{3}\rho_J R^3\,. \tag{4.45}$$

In order to find out what this density is we start off by cubing the expression for the Jeans mass, and then substituting the expression for M_J in terms of ρ_J:

$$M_J^3 = \left(\frac{3k_B T}{2G\bar{m}}\right)^3 R^3 \tag{4.46}$$

$$= \left(\frac{3k_B T}{2G\bar{m}}\right)^3 \frac{3M_J}{4\pi\rho_J} \tag{4.47}$$

$$\rho_J = \frac{3}{4\pi M_J^2}\left(\frac{3k_B T}{2G\bar{m}}\right)^3\,. \tag{4.48}$$

Alternatively, we can write

$$\frac{4\pi}{3}\rho_J R^3 = \frac{3}{2}\frac{k_B T}{G\overline{m}}R \tag{4.49}$$

$$\rho_J = \frac{9}{8\pi}\frac{1}{R^2}\frac{k_B T}{G\overline{m}}\,. \tag{4.50}$$

We will have an instability if the density is larger than this. As we have seen in the previous section, a lower temperature facilitates the collapse. It should be stressed that the precise numerical coefficient will depend on the geometry of the cloud of material, this is not a hard rule but more of a guide for the understanding of the behavior of clouds.

Here appears in the lecture the argument for the fact that the temperature of matter decreases as $T \sim a^{-2}$; it does not really seem to fit with the rest of the chapter, perhaps it should go earlier?
I'll leave it here, commented out.

Equations for stellar structure

In order to properly study the dynamics of the stellar collapse, however, we need to analyze the differential equations which govern it. We will start out by doing so on a static background, following the original reasoning by Jeans (who, working in the early 1900s, did not know about the expansion of the universe). Then, we will discuss the effects of the universe's expansion on the gravitational instability.

The **continuity equation**, imposed by mass conservation, is

$$\partial_t \rho + \nabla \cdot \left(\rho \vec{v} \right) = 0 \,, \tag{4.51}$$

where ρ is the matter density while $\vec{v}$ is the velocity field; the **Euler equation**, imposed by momentum conservation (assuming no viscosity), is

$$\partial_t \vec{v} + \left(\vec{v} \cdot \vec{\nabla} \right) \vec{v} = -\frac{1}{\rho} \vec{\nabla} P - \vec{\nabla} \Phi \,, \tag{4.52}$$

where P is the pressure while Φ is the gravitational potential.

If we define the **convective** time **derivative**,

$$\frac{\mathrm{D}}{\mathrm{D}t} = \partial_t + \vec{v} \cdot \nabla_x = u^\mu \partial_\mu \,, \tag{4.53}$$

we can write the two equations as

$$\frac{\mathrm{D}}{\mathrm{D}t} \rho + \rho \nabla \cdot \vec{v} = 0 \tag{4.54}$$

$$\frac{\mathrm{D}}{\mathrm{D}t} \vec{v} = -\frac{\nabla P}{\rho} - \nabla \Phi \,. \tag{4.55}$$

Lastly, the gravitational field Φ must obey Poisson's equation:

$$\nabla^2 \Phi = 4\pi G \rho \,. \tag{4.56}$$

Right now we have five equations (Euler is a vector equation, corresponding to three scalar ones) and six variables: ρ, Φ, P and the three components of $\vec{v}$. In order to be able to solve this system we need one more condition; typically this is provided as an equation of state, giving P in terms of the other variables.

One way to go about this is to consider entropy: we define the entropy density s by the relation $S = s\rho$, where S is the (total?) entropy.

We will consider isentropic processes, in which

$$\frac{\mathrm{D}s}{\mathrm{D}t} + s \vec{\nabla} \cdot \vec{v} = 0 \,. \tag{4.57}$$

We introduce this, an additional eqution as well as an additional variable, in order to complete our equations with an equation of state in the form:

$$P = P(\rho, s) \,. \tag{4.58}$$

Now, then, we are left with seven equations and seven variables: let us solve them! This is in general very hard, no analytic solutions exist.

Jeans' approach, which we will follow, is to find a fixed background solution and then to perturb it. We are then looking to see whether the perturbation is dampened or amplified. Perturbations are always present, so this will tell us whether the configuration is stable or unstable.

Static ansatz

Jeans' first ansatz was $\rho = \rho_0 = \text{const}$, $\vec{v} = 0$, $s = s_0 = \text{const}$, $\Phi = \Phi_0 = \text{const}$, $P = P_0 = \text{const}$.

This is a *very* simplified model, and it is not even self-consistent: unless $\rho = 0$, Poisson's equation cannot be satisfied, but we want to have matter in our proto-star. We will ignore this problem, since despite it we get a physically meaningful result. The equation cannot precisely hold, but in low-density regions it is not that far from equality.

We perturb the variables: for each variable we will have $x = x_0 + \delta x$ (except $\vec{v}$: since there is no $\vec{v}_0$, we just write $\vec{v}$ instead of $\delta\vec{v}$).

With this, the equations read:

$$\partial_t \delta\rho + \rho_0 \vec{\nabla} \cdot \vec{v} = 0 \tag{4.59}$$

$$\partial_t \vec{v} = -\frac{1}{\rho_0} \vec{\nabla} \delta P - \vec{\nabla} \delta\Phi \tag{4.60}$$

$$\nabla^2 \delta\Phi = 4\pi G \delta\rho \tag{4.61}$$

$$\partial_t \delta s = 0\,. \tag{4.62}$$

The pressure perturbation can be expressed in terms of the density and entropy ones:

$$\delta P = \underbrace{\left.\frac{\partial P}{\partial \rho}\right|_s}_{=c_s^2} \delta\rho + \frac{\partial P}{\partial s} \delta s\,, \tag{4.63}$$

where we recognize the constant-entropy derivative of the pressure with respect to the density: the square of the adiabatic speed of sound.

Next, in order to solve these we need to move to Fourier space. Properly speaking, we would need to take the Fourier transform of all our variables; in terms of computation it is as if we were considering an exponential *ansatz* for all of them, in that taking spatial derivatives corresponds to bringing down a factor $i\vec{k}$:

$$\delta x_i = x_{i0} \exp\left(i\vec{k} \cdot \vec{x} - i\omega t\right)\,, \tag{4.64}$$

with $x_i = \rho, \vec{v}, \Phi, s$.

To first order, then, our equations become:

$$i\omega \delta\rho + \rho_0 i\vec{k} \cdot \vec{v} = 0 \tag{4.65a}$$

$$i\omega \vec{v} = -\frac{1}{\rho_0} i\vec{k} \left(c_s^2 \delta\rho_0 + \frac{\partial P}{\partial s} \delta s \right) - i\vec{k} \delta\Phi_0 \tag{4.65b}$$

$$-k^2 \delta\Phi = 4\pi G \delta\rho \tag{4.65c}$$

$$\omega \delta s = 0\,. \tag{4.65d}$$

Now, the last equation can be solved by either $\omega = 0$ (so, they are time-independent) or $\delta s = 0$ (so, they are isentropic).

These two cases differ substantially in the shape of the vector field they give. A result from Helmholtz is the fact that every velocity field can be decomposed into a divergenceless part and an irrotational part, at least locally: there exist Ψ and $\vec{T}$ such that $\vec{v} = \nabla\Psi + \vec{T}$ with $\nabla \cdot \vec{T} = 0$.

The two cases, as we will see, amount to making the velocity field fully irrotational or fully divergenceless.

Time-independent solutions We start by considering the first option, $\omega = 0$: then, we get

$$\vec{k} \cdot \vec{v} = 0 \tag{4.66a}$$

$$0 = \frac{1}{\rho_0}\vec{k}\left(c_s^2 \delta\rho + \frac{\partial P}{\partial s}\delta s\right) + \vec{k}\delta\Phi \tag{4.66b}$$

$$k^2 \delta\Phi = 4\pi G \delta\rho_0 \,. \tag{4.66c}$$

The first equation tells us that $\vec{k} \cdot \vec{v} = 0$, which in position space translates to $\vec{\nabla} \cdot \vec{v} = 0$, so the velocity field describes the motion of an incompressible fluid.

...which is turbulent? That's not the case in general, water is nearly incompressible but it can have laminar motion. This seems to be what the professor and Pacciani [Pac18b] state though. This is dismissed as uninteresting together with the $\omega = \delta s = 0$ case.

Isentropic solutions Now, we consider the case in which $\delta s_0 = 0$, while $\omega \neq 0$ in general. The equations read

$$\omega\delta\rho + \rho_0 \vec{k} \cdot \vec{v} = 0 \tag{4.67a}$$

$$\omega \vec{v}_0 = \frac{1}{\rho_0}\vec{k} c_s^2 \delta\rho - \vec{k}\delta\Phi \tag{4.67b}$$

$$k^2 \delta\Phi = 4\pi G \delta\rho \,, \tag{4.67c}$$

which we can write as a linear system for the vector $[\delta\rho, \vec{v}, \delta\Phi]$.

The 5x5 coefficient matrix is:

$$\begin{bmatrix} \omega & \rho_0 \vec{k} & 0 \\ \frac{1}{\rho_0}\vec{k}c_s^2 & \omega & \vec{k} \\ 4\pi G & 0 & k^2 \end{bmatrix} , \tag{4.68a}$$

and in order to have more than one solution (we need a family of them, since parameters like ω are variable) we need to set its determinant to zero, which yields:

$$\omega k^2 - \rho_0 \vec{k} \cdot \left(\frac{1}{\rho_0}\vec{k} c_s^2 k^2 - 4\pi G \vec{k}\right) = 0 \,, \tag{4.69}$$

which gives the dispersion relation $\omega^2 = c_s^2 k^2 - \rho_0 4\pi G$.

This has a direct physical interpretation: if ω^2 is positive, then ω is real so the solution is oscillatory; while if ω^2 is negative then ω is imaginary, therefore the solution is given by a *real* exponential, which quickly amplifies (or dampens, but that case is not interesting since it does not have macroscopic effects) the perturbation. This is the *unstable* case.

A generic solution will be a combination of these, with varying $\vec{k}$.

We can connect this result with what we have found for the freefall timescale earlier: if the pressure is negligible then so is c_s, so we have $\omega^2 = -\rho_0 4\pi G$: so, we have a solution increasing on a timescale dictated by $|\omega| = (4\pi G\rho_0)^{1/2}$, the characteristic time

$$\tau = \frac{1}{|\omega|} = \frac{1}{\sqrt{4\pi G\rho_0}}\,, \tag{4.70}$$

which is very similar to the freefall timescale

$$\tau_{\text{free fall}} = \left(\frac{3\pi}{32G\rho_0}\right)^{1/2}; \tag{4.71}$$

the difference is only the numerical factor in front, and they are quite similar (0.28 versus 0.54).

The separation between real and imaginary ω is reached when $\omega^2 = 0$, which gives us the critical *Jeans wavenumber*:

$$k_J^2 = \frac{4\pi G\rho_0}{c_s^2}\,. \tag{4.72}$$

The wavenumber k_J also defines a wavelength $\lambda_J = 2\pi/k_J$, which will then tell us what the length scale above which we have instability is.

This is, up to a order-1 difference in the numerical factor in front, the same result we had before: we can see this if we consider the fact that, for an ideal gas, $c_s \approx \sqrt{k_B T/\overline{m}}$ we recover

$$\lambda_J \sim \frac{1}{\sqrt{G\rho_0}}\sqrt{\frac{k_B T}{\overline{m}}}\,, \tag{4.73}$$

which is consistent with equation (4.50).

A plasma analogy The result we found is similar to what we get with a plasma of charged particles, with the electrostatic potential instead of the gravitational field. In that case, the dispersion relation is given by

$$\omega^2 = c_s^2 k^2 + \frac{4\pi n_e e^2}{m_e}\,, \tag{4.74}$$

where n_e is the number density of electrons, m_e is the electron mass.

We have the following analogies:

$$n_e \to \rho_0/m \tag{4.75a}$$

$$m_e \to m \tag{4.75b}$$

$$e^2 \to Gm^2\,. \tag{4.75c}$$

The equations are similar but, crucially, the sign of the additional term in the dispersion relation is positive in the gravitational case and negative in the electromagnetic case. This is due to the fact that there exists only positive gravitational "charge", while we have both positive and negative electric charge: in the electromagnetic case, then, we can have screening effects, not so in the gravitational one.

Expanding ansatz

Now, we will study the same problem, but instead of a static background we will use an expanding one, described by a flat FLRW metric (the effects of spatial curvature will be negligible on the scale of a stellar formation cloud anyway).

The physical radial vector will be given by $\vec{r} = a(t)\vec{x}$, where $\vec{x}$ is the radial vector in comoving coordinates.

We now drop the vector sign, but still imply it; the physical velocity is given by

$$u = \dot{r} = \dot{a}x + a\dot{x} = \frac{\dot{a}}{a}r + v = Hr + v\,, \tag{4.76}$$

where $v = a\dot{x}$ is called the **peculiar velocity**. This has a direct physical implication: the distant galaxies we observe are, generally speaking, the more redshifted the further they are from us; however, there is noise in this relation due to the Doppler shift caused by the peculiar velocity. An extreme example of this is the Andromeda galaxy which, despite being almost a Mpc away, is actually *blueshifted* to $z \approx -0.001$ since its peculiar velocity is directed towards the Solar System.

Let us then seek perturbed solutions to the equations of motion of the fluid. We will neglect the pressure gradient in order to simplify the considerations — this is not a great approximation, since we are always perturbing around the equilibrium situation in which the instability has not yet formed, so in a low-pressure scenario. When the pressure will start to take over we will be far from the background anyway, so we cannot hope to describe the situation in this way. This discussion will then also apply to dark-matter structure formation, since it is pressureless.

We will use a slightly different notation from before, denoting the background with an index b, so that

$$\rho(\vec{r},t) = \rho_b(t) + \delta\rho(\vec{r},t) \tag{4.77}$$

$$\vec{v}(\vec{r},t) = \vec{v}_b(\vec{r},t) + \vec{v}(\vec{r},t) \tag{4.78}$$

$$\Phi(\vec{r},t) = \Phi_b(\vec{r},t) + \phi(\vec{r},t)\,. \tag{4.79}$$

The equations, using the customary notation for a partial derivative taken while keeping a certain variable constant, read:

$$\left.\frac{\partial \rho}{\partial t}\right|_{\vec{r}} + \nabla_{\vec{r}} \cdot \left(\rho\vec{u}\right) = 0 \tag{4.80}$$

$$\left.\frac{\partial \vec{u}}{\partial r}\right|_{\vec{r}} + (\vec{u}\cdot\nabla_{\vec{r}})\vec{u} = -\nabla_{\vec{r}}\Phi \tag{4.81}$$

$$\nabla^2_{\vec{r}}\Phi = 4\pi G\rho\,. \tag{4.82}$$

Let us now try to find the **background** solution. We assumed that the background density of matter $\rho_b(t)$ is space independent but time dependent: therefore, up to a constant (which we set to zero) and a linear term in $\vec{r}$ (which would violate isotropy) we must

$$\Phi_b(\vec{r},t) = \frac{2\pi G}{3}\rho_b(t)r^2\,, \tag{4.83}$$

which is consistent with what we found earlier, since its gradient and then Laplacian are:

$$\nabla_{\vec{r}}\Phi_b = \frac{4\pi G}{3}\rho_b(t)\vec{r} \tag{4.84}$$

$$\nabla^2_{\vec{r}}\Phi_b = 4\pi G\rho_b\,. \tag{4.85}$$

This potential diverges at $r\to\infty$; however this is not an issue with the solution, but with the Newtonian approximation we made. It will not affect our treatment of the problem.

We want to get equations in the comoving coordinates ($\vec{x}$), not the local inertial ones ($\vec{r}$). Let us consider a generic function $f(\vec{r},t)$, which can also be expressed with respect to $(\vec{x},t)$.

It can be shown[4] that the difference between the time derivatives at fixed $\vec{r}$ and at fixed $\vec{x}$ of a generic function f (which can also be a vector) is given by

$$\left.\frac{\partial f}{\partial t}\right|_{\vec{x}} = \left.\frac{\partial f}{\partial t}\right|_{\vec{r}} + H(\vec{r}\cdot\nabla_{\vec{r}})f\,. \tag{4.88}$$

We can then mold the continuity equation:

$$\left.\frac{\mathrm{d}\rho}{\mathrm{d}t}\right|_{\vec{x}} - H(\vec{r}\cdot\nabla_{\vec{r}})\rho + \rho\vec{\nabla}\cdot\vec{u} + (\vec{u}\cdot\nabla_{\vec{r}})\rho = 0 \tag{4.89}$$

$$\left.\frac{\mathrm{d}\rho}{\mathrm{d}t}\right|_{\vec{x}} - H(\vec{r}\cdot\nabla_{\vec{r}})\rho + \rho\vec{\nabla}_{\vec{r}}\cdot(H\vec{r}+\vec{v}) + H\left(\vec{r}\cdot\vec{\nabla}_{\vec{r}}\right)\rho + \left(\vec{v}\cdot\vec{\nabla}_{\vec{r}}\right)\rho = 0 \tag{4.90}$$

$$\left.\frac{\mathrm{d}\rho}{\mathrm{d}t}\right|_{\vec{x}} + 3H\rho + \rho\vec{\nabla}_{\vec{r}}\cdot\vec{v} + \left(\vec{v}\cdot\vec{\nabla}_{\vec{r}}\right)\rho = 0 \tag{4.91}$$

$$\left.\frac{\partial\rho}{\partial t}\right|_{\vec{x}} + 3H\rho + \underbrace{\frac{1}{a}\nabla_{\vec{x}}}_{\vec{\nabla}_{\vec{r}}}\cdot(\rho\vec{v}) = 0\,. \tag{4.92}$$

[4] The way to go about it is to impose the equality of the total time derivatives of $f(\vec{x},t)$ and $f(\vec{r},t)$, which read:

$$\frac{\mathrm{D}f(\vec{r},t)}{\mathrm{D}t} = \frac{\partial f}{\partial t}_{\vec{r}} + (\nabla_{\vec{r}}f)\cdot\underbrace{(\vec{v}+H\vec{r})}_{\dot{r}} \tag{4.86}$$

$$\frac{\mathrm{D}f(\vec{x},t)}{\mathrm{D}t} = \frac{\partial f}{\partial t}_{\vec{x}} + (\nabla_{\vec{x}}f)\cdot\underbrace{\vec{v}}_{\dot{x}}\,. \tag{4.87}$$

As we would expect, if the velocity $\vec{v}$ is equal to zero then the density scales like $\rho \sim a^{-3}$.

The computation for the Euler equation is an application of the same principles, with one exception: we can simplify the background potential term with some terms which appear on the left-hand side, using the equation

$$\left.\frac{\partial(H\vec{r})}{\partial t}\right|_{\vec{r}} + H(\vec{r}\cdot\nabla_{\vec{r}})(H\vec{r}) = -\nabla_{\vec{r}}\Phi_b\,. \tag{4.93}$$

The divergence term is just $r^j\partial_j r^i = r^i$. Inserting the expression we know for the background gravitational potential, whose gradient is proportional to $\vec{r}$, we get:

$$\vec{r}\left(\dot{H} + H^2 = -\frac{4\pi G}{3}\rho_b\right), \tag{4.94}$$

which must hold for any $\vec{r}$, so we drop it and recover the second Friedmann equation (using $\dot{H} + H^2 = \ddot{a}/a$). Working backwards, we can prove the equation.

Using this, we can write the Euler equation as:

$$\left.\frac{\partial\vec{v}}{\partial t}\right|_{\vec{x}} + H\vec{v} + \frac{1}{a}(\vec{v}\cdot\nabla_{\vec{x}})\vec{v} = -\frac{1}{a}\nabla_{\vec{x}}\delta\Phi\,; \tag{4.95}$$

getting the Poisson equation in comoving coordinates is faster and directly yields

$$\nabla^2\delta\Phi = a^2 4\pi G\delta\rho\,. \tag{4.96}$$

We express the density perturbation by defining the variable δ as

$$\delta(\vec{x},t) = \frac{\delta\rho}{\rho_b} = \frac{\rho(\vec{x},t) - \rho_b(t)}{\rho_b(t)}\,, \tag{4.97}$$

which can take values anywhere from -1 to $+\infty$: we can interpret a negative δ as a sort of "negative effective gravitational charge". This has the physical meaning that we can expect screenings: under-dense and over-dense regions can "balance out" at large distances, just like we do not observe the effects of large-scale electric charge imbalances.

We can then give a quantitative check of whether the Newtonian approximation is a good one. In terms of δ, the Poisson equation reads:

$$\nabla^2\delta\Phi = 4\pi G\rho_b\delta a^2\,, \tag{4.98}$$

which we can estimate: let us say that λ is the typical variation scale of the potential, so that $\nabla^2\delta\Phi \sim \lambda^{-2}\delta\Phi$; also let us use the first Friedmann equation relating H^2 with the background density ρ_b (since the FE only hold at large scales):

$$\lambda^{-2}\delta\Phi \sim 4\pi G\frac{H^2}{8\pi G/3}a^2\delta \tag{4.99}$$

$$\delta\Phi \sim \frac{3}{2}H^2\delta a^2\lambda^2 \sim \left(\frac{\lambda^2}{\lambda_{\text{hor}}^2}\right)\delta\,, \tag{4.100}$$

where the typical variation of the gravitational field is $\lambda \sim$ Mpc, while $c/(Ha) = \lambda_{\text{hor}} \sim$ Gpc is the (comoving) Hubble horizon scale.

The density perturbation will be, at most, of order 1, so we get that $\delta\Phi$ is indeed small. As long as the perturbations are only galactic, the Newtonian approximation is good.

Solving the equations The derivative of the density, which appears in the continuity equation, reads:

$$\frac{\partial \rho}{\partial t} = \frac{\partial \rho_b}{\partial t}(1+\delta) + \rho_b \frac{\partial \delta}{\partial t} \,. \tag{4.101}$$

The procedure we will then apply is to simplify the equation making use of the fact that it holds at zeroth order (with all the perturbations set to zero): we have $\partial_t \rho_b + 3H\rho_b = 0$.

Also, we neglect second and higher order terms (recall that $\vec{v}$ is already first order): the computation goes like

$$\frac{\partial \rho_b}{\partial t}(1+\delta) + \rho_b \frac{\partial \delta}{\partial t} + 3H\rho_b(1+\delta) + \frac{1}{a}\vec{\nabla}\cdot(\rho_b(1+\delta)v) = 0 \tag{4.102}$$

$$(1+\delta)\left[\frac{\partial \rho_b}{\partial t} + 3H\rho_b\right] + \rho_b \frac{\partial \delta}{\partial t} + \frac{\rho_b}{a}\vec{\nabla}\cdot\vec{v} = 0 \tag{4.103}$$

$$\frac{\partial \delta}{\partial t} + \frac{1}{a}\vec{\nabla}\cdot\vec{v} = 0 \,. \tag{4.104}$$

On the other hand, for the momentum equation we only need to neglect the $(\vec{v}\cdot\vec{\nabla})\vec{v}$ term, which is second order, to find:

$$\frac{\partial \vec{v}}{\partial t} + H\vec{v} = -\frac{1}{a}\vec{\nabla}\phi \,, \tag{4.105}$$

where $\phi = \delta\Phi$ is the perturbation

Finally, the Poisson equation is already in its simplest, linear form.

In order to solve these we expand in Fourier space: the density perturbation δ is expressed in terms of its Fourier transform $\widetilde{\delta}$ as

$$\delta(\vec{x},t) = \frac{1}{(2\pi)^3}\int \mathrm{d}^3\vec{k}\,\widetilde{\delta}(t)\exp\Big(i\vec{k}\cdot\vec{x}\Big) \,, \tag{4.106}$$

and for $\vec{v}$ and ϕ. Note that we only expand in 3D space: plane waves do not propagate nicely in an expanding universe, so they are not a good Fourier base: we keep time derivatives, substituting spatial ones with multiplication by $i\vec{k}$.

The actual quantities must be real: therefore we know that $\widetilde{\delta}^*_{\vec{k}}(t) = \widetilde{\delta}_{-\vec{k}}(t)$.

Before we Fourier-transform the equations, we can make a **simplification**: as we mentioned before, any vector field $\vec{v}$ can be decomposed by the Helmholtz theorem into the gradient of a scalar field and a divergenceless vector field:

$$\vec{v} = \nabla\Psi + \vec{T} \,, \tag{4.107}$$

where $\nabla\cdot\vec{T} = 0$ (and, as is known from vector calculus, $\nabla\times(\nabla\Psi) = 0$).

Before we considered only the first order, the Euler equation read:

$$\frac{\mathrm{D}\vec{v}}{\mathrm{D}t} + H\vec{v} = -\frac{1}{a}\nabla\phi \,. \tag{4.108}$$

If we substitute $\vec{v}$ with its Helmholtz decomposition, we can split the equation into two, one for $\vec{T}$ and one for Ψ:

$$\frac{D\vec{T}}{Dt} + H\vec{T} = 0 \qquad \text{and} \qquad \frac{D(\vec{\nabla}\Psi)}{Dt} + H\vec{\nabla}\Psi = -\frac{\vec{\nabla}\phi}{a}. \tag{4.109}$$

We could determine which terms went on either side based on whether they had zero divergence or zero curl.

Does the convective derivative commute with $\nabla\times$ and $\nabla\cdot$, though?

This has an important physical meaning: the divergenceless part of the velocity evolves by itself, unaffected by the gravitational field perturbation. The way it evolves is, roughly speaking, a decreasing exponential, so its magnitude will diminish over time. Therefore, any $\vec{T}$ component which is part of the velocity field initially gets ever more diluted.

Because of this, we neglect the divergenceless component of the velocity field, and only consider the $\vec{v} = \nabla\Psi$ part. In Fourier space, this reads $\vec{v} \propto \vec{k}\Psi$, so we can project the three equations along the unit vector $\hat{k} = \vec{k}/\left|\vec{k}\right|$, simplifying them to a single scalar one. We will denote $v = \vec{v}\cdot\hat{k}$ and $k = \vec{k}\cdot\hat{k} = \left|\vec{k}\right|$.

With this and denoting time derivatives with a dot, our equations will read:

$$\dot{\delta} + \frac{ik}{a}v = 0 \tag{4.110}$$

$$\dot{v} + Hv = -\frac{ik}{a}\phi \tag{4.111}$$

$$-k^2\phi = 4\pi G a^2 \rho_b \delta. \tag{4.112}$$

We can find an equation for δ alone by differentiating the first equation with respect to time and substituting the Euler equation:

$$\ddot{\delta} + \frac{ik}{a}\dot{v} - \frac{ik}{a^2}\dot{a}v = 0 \tag{4.113a}$$

$$\ddot{\delta} + \frac{ik}{a}\left(-Hv - \frac{ik}{a}\phi\right) - \frac{ik}{a}Hv = 0 \tag{4.113b}$$

$$\ddot{\delta} - \frac{2ik}{a}Hv + \frac{k^2\phi}{a^2} = 0, \tag{4.113c}$$

but, from the Poisson equation, the last term is equal to $-4\pi G\rho_b\delta$, and from the continuity equation again the second term is equal to $2H\dot{\delta}$: the equation then reads

$$\ddot{\delta} + 2H\dot{\delta} - 4\pi G\rho_b\delta = 0. \tag{4.114}$$

This looks promising: if H and ρ_b were constant, it would be a simple second-order ODE. They are not constant, but they are functions corresponding to the background: we can use the solutions found earlier corresponding to a matter-dominated universe,

$$a(t) \propto t^{2/3}, \qquad H = \frac{2}{3t}, \qquad \rho_b = \left(6\pi G t^2\right)^{-1}. \tag{4.115}$$

With these, the equation reads

$$\ddot{\delta} + \frac{4}{3t}\dot{\delta} - \frac{2}{3t^2}\delta = 0\,, \tag{4.116}$$

which will have two independent solutions since it is of second order: it turns out that both can be recovered using a powerlaw ansatz, $\delta \propto t^\alpha$: the equation for α reads

$$\alpha(\alpha - 1) + \frac{4}{3}\alpha - \frac{2}{3} = 0\,, \tag{4.117}$$

whose solutions are $\alpha = -1$ and $\alpha = 2/3$.

The solution with $\alpha = 2/3$ is the **growing mode**, while the one with $\alpha = -1$ (so, $\delta \propto t^{-1} \propto H(t)$) is the **decaying mode**.

1. The growing mode has $\delta \propto t^{2/3} \propto a(t)$, $v \propto t^{1/3}$, $\phi = \text{const}$;
2. the decaying mode has $\delta \propto t^{-1} \propto H(t)$, $v \propto t^{-1/2}$, $\phi \propto t^{-5/3}$.

Typically, we are more interested in the growing mode.

"When the inflaton field becomes classical we lose a degree of freedom: this removes the decaying solutions. ": not really clear to me.

4.2.1 Star formation

Let us now actually discuss stellar formation specifically. The presence of molecular hydrogen, H_2, is correlated to stellar formation since it can absorb some of the kinetic energy of the collapsing cloud by dissociating, thus allowing for further collapse; it forms through the channels

$$\mathrm{H} + e \leftrightarrow \mathrm{H}^- + \gamma \qquad \mathrm{H}^- + \mathrm{H} \leftrightarrow \mathrm{H}_2 + e \tag{4.118a}$$

$$\mathrm{H} + p \leftrightarrow \mathrm{H_2}^+ + \gamma \qquad \mathrm{H_2}^+ + \mathrm{H} \leftrightarrow \mathrm{H}_2 + \mathrm{p}\,, \tag{4.118b}$$

and these processes, generally speaking, start to become efficient at redshifts of about $z \sim 200$.

As we have seen, if we fix the temperature then the Jeans critical density scales like $\rho_J \propto M^{-2}$ (4.48): so, it is easier to get above the Jeans density if the mass is high (and the temperature is low). This means that we expect the formation of stars to be a top-down process: larger structures form first.

Let us now give some quantitative estimates of the densities at hand.

Approximately, the baryonic density today is $\rho_{0b} \sim 1 \times 10^{-28}\,\mathrm{kgm^{-3}}$.

Going backwards in time, the baryonic density scales like $\rho_b(z) = (1+z)^3 \rho_{0b}$: at $z \sim 200$, for example, we have $\rho_b(z \sim 200) \sim 10^{-22}\,\mathrm{kgm^{-3}}$.

Stars will not form anywhere, they will only do so in over-dense regions, which form preferentially in the centers of dark-matter halos, acting as gravitational "traps".

Let us take, as an example, a molecular hydrogen ($\overline{m} \approx 2m_p$) cloud with a mass of $M = 1000 M_\odot$ and $T \approx 20\,\mathrm{K}$ (a typical temperature for matter at this redshift): we have a Jeans critical density of $\rho_J \approx 4 \times 10^{-22}\,\mathrm{kg m^{-3}}$.

Perhaps this is the place to put the scaling of the temperature of matter! The fact that the temperature of baryons decreases slower than the radiation's is crucial to reach a temperature low enough to reach the Jeans bound.

For a solar-mass cloud, the critical density is much lower: $\rho_J \approx 4 \times 10^{-16}\,\mathrm{kg m^{-3}}$.

4.2.2 Collapsing a solar-mass cloud

Let us suppose that, by virtue of being in a dark matter halo and after some large-scale collapse, we have indeed reached the conditions of $\rho \sim 4 \times 10^{-16}\,\mathrm{kg m^{-3}}$ and $T \sim 20\,\mathrm{K}$. We will try to understand how a Sun-like star might form.

Molecular hydrogen is present in the cloud, its dissociation energy is $\epsilon_D \approx 4.5\,\mathrm{eV}$ while, as we know, the ionization energy of hydrogen is $\epsilon_H \approx 13.6\,\mathrm{eV}$.

As the collapse starts, any kinetic energy which is developed is used, first to dissociate molecular hydrogen and then to ionize hydrogen, so for a while the evolution goes according to the free-fall equation of motion:

$$\frac{1}{2}\left(\frac{\mathrm{d}r}{\mathrm{d}t}\right)^2 = \frac{Gm_0}{r} - \frac{Gm_0}{r_0}\,, \tag{4.119}$$

which, as we have seen earlier, corresponds to a free-fall time given by (4.13): we get $t_{FF} \approx 100\,\mathrm{kyr}$.

The energy needed to dissociate all the H_2 and then ionize the H is given in terms of the mass of the cloud, $M = M_\odot$, by

$$E = \frac{M}{2m_H}\epsilon_D + \frac{M}{m_H}\epsilon_I\,, \tag{4.120}$$

where $m_H \approx m_p$ is the mass of hydrogen. The mass of H_2 is slightly different from $2m_H$, but the difference is of the order of $\epsilon_D / m_H \sim 10^{-8}$, completely negligible.

The cloud starts from a radius $R_1 \sim \sqrt[3]{3M/4\pi\rho_J} \approx 1 \times 10^{15}\,\mathrm{m} \approx 0.1\,\mathrm{lyr}$: what radius R_2 does it reach by the time all the hydrogen is ionized?

We can calculate it by equating the difference in potential energy to the ionization and dissociation energy:

$$GM_\odot^2\left(\frac{1}{R_2} - \frac{1}{R_1}\right) = E = \frac{M_\odot}{2m_H}\epsilon_D + \frac{M_\odot}{m_H}\epsilon_I \tag{4.121}$$

$$GM_\odot\left(\frac{1}{R_2} - \frac{1}{R_1}\right) = \frac{\epsilon_D}{2m_H} + \frac{\epsilon_I}{m_H} \approx 1.7 \times 10^{-8} c^2 \tag{4.122}$$

$$\left(\frac{1}{R_2} - \frac{1}{R_1}\right)^{-1} \approx \frac{GM_\odot}{c^2} \times 6 \times 10^7 \approx 9 \times 10^{10}\,\mathrm{m}\,, \tag{4.123}$$

so, since $R_1 \gg 10^{11}\,\mathrm{m}$ the R_1^{-1} term is basically negligible, while $R_2 \approx 10^{11}\,\mathrm{m}$. We have shrunk our cloud from 0.1 lyr to about 0.6 AU, just smaller than the radius of the orbit of Venus. This is still much larger than $R_\odot \approx 7 \times 10^8\,\mathrm{m}$.

The rest of the collapse is much slower, and it happens under hydrostatic equilibrium; the proto-star will still shrink under its own gravity, getting ever hotter, until it reaches the temperature needed for the ignition of fusion, around 10^7 K.

At this stage the virial theorem applies, since the plasma is optically thick, very little energy is lost to radiation:

$$2E_k + E_{\mathrm{gr}} = 0\,, \tag{4.124}$$

and we can recover the total kinetic energy from the equipartition theorem:

$$E_k = \frac{3}{2} N k_B T = \frac{3}{2} \frac{M}{\overline{m}} k_B T \approx \frac{3M}{m_H} k_B T\,, \tag{4.125}$$

where we used the fact that $\overline{m} = 0.5 m_H$: the hydrogen is ionized, so we have both free electrons and free protons, and basically all the mass is with the latter.

The gravitational binding energy can be estimated as what was lost in the dissociation and ionization:

$$2 \times \underbrace{3k_B T \frac{M}{m_H}}_{E_k} = -E_{\mathrm{gr}} \approx \frac{M}{m_H} \left(\frac{\epsilon_D}{2} + \epsilon_I \right) \tag{4.126}$$

$$k_B T \approx \frac{1}{6} \left(\frac{\epsilon_D}{2} + \epsilon_I \right) \approx 2.6\,\mathrm{eV} \approx 30\,\mathrm{kK}\,, \tag{4.127}$$

still very much lower than the temperature needed to ignite fusion, which is on the order of a keV $\sim 10\,$MK.

4.2.3 Conditions for stardom and brown dwarfs

At this point, the proto-star reaching a high enough temperature to fuse hydrogen is not a given: it may not happen.

In order to discuss the *conditions for stardom* we need to account (albeit in an approximate way, as usual) for the fermionic nature of protons and electrons, which will give us a maximum density due to the Pauli exclusion principle: the particles will form a degenerate Fermi gas. Depending on the mass of the star, this maximum density may be reached before the ignition temperature: when this is the case, a **brown dwarf** is formed.

We estimate the minumum space a fermion can occupy with its De Broglie wavelength:

$$\lambda = \frac{h}{p} = \frac{2\pi\hbar}{p}\,, \tag{4.128}$$

where p is the momentum of the particle. Since the temperature is still less than a keV, both electrons and protons are nonrelativistic, so we can approximate $E_k = p^2/2m$, which will typically be of the order $E_K \sim k_B T$.

Then, the momentum and wavelength will be:

$$p \sim \sqrt{2mk_B T} \implies \lambda \sim \frac{2\pi\hbar}{\sqrt{mk_B T}}\,. \tag{4.129}$$

We can see that $\lambda \sim 1/\sqrt{m}$: this means that the wavelength is smaller (by a factor 40) for the protons than it is for the electrons, and they appear in equal numbers; therefore the first bound to be reached will be that of the degenerate electron gas, for this reason we will neglect protons now.

The critical density will be reached when we have an electron for each λ^3; however in order to ensure local charge neutrality we must have protons distributed in the same way, and their mass contribution to the density will be the largest:

I'd actually say, then, that we should write $\rho_c = m_p/\lambda^3$!

$$\rho_c = \frac{\overline{m}}{\lambda^3} \sim \overline{m}\frac{(m_e k_B T)^{3/2}}{(2\pi\hbar)^3}\,. \tag{4.130}$$

As we have seen earlier, since the virial theorem applies the temperature of the proto-star is tied to its gravitational binding energy by

$$3Nk_B T = \frac{GM^2}{R} \implies k_B T = \frac{GM\overline{m}}{3R}\,. \tag{4.131}$$

The mass can be expressed in terms of the average density as $M = \frac{4}{3}\pi\overline{\rho}R^3$, so

$$\frac{1}{R} = \left(\frac{4\pi}{3}\frac{\overline{\rho}}{M}\right)^{1/3}, \tag{4.132}$$

which gives us the result

$$k_B T = \frac{GM\overline{m}}{3}\left(\frac{4\pi}{3}\frac{\overline{\rho}}{M}\right)^{1/3}. \tag{4.133}$$

If we substitute the critical density ρ_c for $\overline{\rho}$ we will get the maximum possible temperature allowed at a given mass M: after some manipulation we get (up to a small constant, which we neglect since the calculation is rough anyway):

$$k_B T = \frac{GM\overline{m}}{3}\left(\frac{4\pi}{3M}\right)^{1/3}\frac{\overline{m}^{1/3}}{(2\pi\hbar)}(m_e k_B T)^{1/2} \tag{4.134}$$

$$(k_B T)^2 = \frac{G^2M^2\overline{m}^2}{9}\left(\frac{4\pi}{3M}\right)^{2/3}\frac{\overline{m}^{2/3}}{(2\pi\hbar)^2}m_e k_B T \tag{4.135}$$

$$k_B T \approx \frac{G^2\overline{m}^{8/3}M^{4/3}m_e}{(2\pi\hbar)^2}\,. \tag{4.136}$$

Inserting the ignition temperature of around 1 keV we get $M_{\text{min}} \approx M_\odot$, which is almost the right order of magnitude: more accurate models agree with the observational result of $M_{\text{min}} \approx 0.08 M_\odot$.

4.3 The Sun and other stars

Let us start off with a list of the physical characteristics of the Sun: its mass is $M_\odot \approx 1.99 \times 10^{30}$ kg, its radius is $R_\odot \approx 6.96 \times 10^8$ m, its bolometric luminosity[5] is $L_\odot = 3.86 \times 10^{26}$ W.

[5] This means: the total luminosity, integrated across all the electromagnetic spectrum.

Its age is around $t_\odot \approx 4.55 \times 10^9\,\mathrm{yr}$, which is comparable to the age of the Universe.

At its core, the density is $\rho_c \approx 1.48 \times 10^5\,\mathrm{kgm^{-3}}$, the temperature is $T_c = 1.56 \times 10^7\,\mathrm{K} \approx 1.3\,\mathrm{keV}$, and the pressure is around $P_c = 2.29 \times 10^{16}\,\mathrm{Pa}$.

The radiation emitted by the Sun approximately follows a blackbody curve, whose characteristic temperature is called the *effective temperature*: it is around $T_E \approx 5780\,\mathrm{K} \approx 0.5\,\mathrm{eV}$.

The power emitted by the Sun is quite small as a fraction of its total energy, so we can apply the virial theorem: one of its formulations we found is

$$\langle P \rangle = -\frac{1}{3}\frac{E_{\mathrm{gr}}}{V}\,, \tag{4.137}$$

where $E_{\mathrm{gr}} \approx -GM^2/R$ while $V = 4\pi R^3/3$: plugging the Sun's numbers we get

$$\langle P \rangle = \frac{GM_\odot^2}{4\pi R_\odot^4} \approx 10^{14}\,\mathrm{Pa}\,, \tag{4.138}$$

100 times less than the central pressure. Similarly, the average density can be computed as $\langle \rho \rangle = M_\odot/V \approx 1.4 \times 10^3\,\mathrm{kg/m^3}$, just slightly more than the density of water! This is also roughly 100 times less than the central value.

By the ideal gas law (which does not precisely apply, but as we saw the Sun has a rather low density overall so it is a reasonable approximation) we can find the mean internal temperature of the Sun, T_I, as:

$$\langle P \rangle = \frac{\langle \rho \rangle}{\overline{m}} k_B T_I\,, \tag{4.139}$$

where $\overline{m}$, the average particle mass, is $\overline{m} \approx 0.61 m_H$ instead of $0.5 m_H$ when considering the proper composition of the Sun, which has a noticeable fraction of helium and heavier elements.

We get

$$k_B T_I = \frac{GM_\odot^2}{4\pi R_\odot^4}\left(\frac{M_\odot}{4\pi R_\odot^3/3}\right)^{-1}\overline{m} = \frac{GM_\odot \overline{m}}{3R_\odot} \approx 0.5\,\mathrm{keV} \approx 5 \times 10^6\,\mathrm{K}\,. \tag{4.140}$$

Again, we see the relation $x_{\mathrm{central}} \ll x_{\mathrm{mean}} \ll x_{\mathrm{surface}}$, which holds for $x =$ density, pressure, temperature.

The measured bolometric luminosity of the Sun is consistent with

$$L_\odot = 4\pi R_\odot^2 \sigma T_E^4\,, \tag{4.141}$$

where σ is Stefan's constant: $\sigma \approx 5.67 \times 10^{-8}\,\mathrm{Wm^{-2}K^{-4}}$.

Why does this hold with the comparatively low temperature T_E and not with the mean internal temperature T_I? We will now see that this is because the interior of the Sun is optically thick (opaque), so any photon from the interior cannot simply be emitted, it will undergo many scatterings before doing so.

4.3.1 Radiative diffusion

The motion of the photon is Brownian, and the total displacement $\vec{D}$ from production to emission will be written in terms of N short straight tracts,

$$\vec{D} = \sum_{i=1}^{N} \vec{\ell}_i \,. \tag{4.142}$$

The process is inherently stochastic, and we will need to describe it as such. We need two equations in order to describe it: the Langevin and Fokker-Planck equations.

The **Langevin** equation gives us the derivative of the position of the particle in terms of a stochastic force η:

$$\dot{\vec{D}}(t) = \vec{\eta}(t) \,. \tag{4.143}$$

The requirements on η are that it should have zero mean: $\left\langle \vec{\eta} \right\rangle(t) = 0$ and that it should satisfy $\left\langle \eta_i(t)\eta_j(t') \right\rangle = 2D\delta_{ij}\delta(t-t')$. This means that it is spatially and temporally uncorrelated, making the process a Markovian one: it has "no memory". The parameter D is called the *diffusion constant*, its units are m^2/s (and it is not related to the displacement $\vec{D}$).s

Nontrivially (we will not discuss the details of the derivation) this gives us the **Fokker-Planck** formula:

$$\frac{\partial P}{\partial t} = D\nabla^2 P \,, \tag{4.144}$$

where P, a function of position and time, quantifies the probability density of finding a photon there.

Mathematically speaking this is a *parabolic differential equation*, which concretely means that we need to give it both initial conditions and boundary conditions.

The kind of boundary condition we want for the study of the Sun is a so-called *absorbing boundary*: physically, as a photon reaches the surface it is emitted, which from the inside looks like if the boundary "absorbed it".

The solution to the Fokker-Planck equation, with an impulsive initial condition like $P(x, t=0) = \delta^{(3)}(x)$ is a Gaussian with variance $\left\langle x^2 \right\rangle = 2Dt$ and centered around zero:

$$P(x,t) = \frac{1}{(4\pi Dt)^{3/2}} \exp\left(-\frac{x^2}{4Dt}\right) . \tag{4.145}$$

This holds for $0 < |\vec{x}| < R_\odot$, while for $|x| > R_\odot$ the particles have escaped.

Integrating the Gaussian within the stellar boundary at a time t yields the probability that any single photon emitted at the center of the Sun is still inside it after a time t. Roughly, this be the case with $\sigma = \sqrt{2Dt} \lesssim R_\odot$.

Let us now consider the problem geometrically. The mean square value of the displacement will be given by

$$\left\langle \vec{D}^2 \right\rangle = \sum_i \left\langle \vec{\ell}_i^2 \right\rangle + 2\sum_{i<j} \left\langle \vec{\ell}_i \cdot \vec{\ell}_j \right\rangle \,, \tag{4.146}$$

but if we have isotropy then the scalar products have mean zero, since they are averages of two lengths times a cosine.

Then we find that, in order for the photon to have reached the boundary of the star on average it will need to have undergone N scatterings: if we set $\left\langle \vec{D}^2 \right\rangle = R_\odot^2$ we get

$$\left\langle \vec{D}^2 \right\rangle = N\ell^2 = R_\odot^2 \,, \tag{4.147}$$

so $N = R_\odot^2/\ell^2$, where ℓ is the typical path travelled between scatterings.

The time it takes for a photon to cover a distance ℓ is $t = \ell/c$. Then, the total time taken in the random walk is given by $t_{RW} = Nt = R_\odot^2 \ell/(\ell^2 c)$, which means

$$t_{RW} = \frac{R_\odot^2}{c\ell} \,, \tag{4.148}$$

while in direct flight the photon would only have taken $t_0 = R_\odot/c$: their ratio is

$$\frac{t_{RW}}{t_0} = \frac{R_\odot}{\ell} \,. \tag{4.149}$$

The argument which follows is still unconvincing to me: the photons take $\sim 50\,\text{kyr}$ to come out of the Sun, so this process will have stabilized in the Sun's $5\,\text{Gyr}$ lifetime! Estimating the mean free path as $\ell \sim 1/(n\sigma_T)$ could work, it yields $\sim 1\,\text{cm}$ with the mean density and $\sim 0.1\,\text{mm}$ with the central density, so perhaps since in the low-ℓ regions the photons stay for a longer time this can average out to $1\,\text{mm}$ when doing the calculation properly.

If $L'_\odot$ is the luminosity for the Sun calculated using the average internal temperature T_I instead of the effective temperature T_E we get

$$L_\odot = L'_\odot \frac{\ell}{R_\odot} \,, \tag{4.150}$$

which means

$$T_E = \left(\frac{\ell}{R_\odot} \right)^{1/4} T_I \,, \tag{4.151}$$

so we can calculate ℓ by knowing the other three parameters: we get that the mean free path, averaged over the star, is $\ell \approx 1\,\text{mm}$.

The total solar luminosity is then:

$$L_\odot = L'_\odot \frac{\ell}{R} = 4\pi R_\odot^2 \sigma T_I^4 \frac{\ell}{R} \,, \tag{4.152}$$

and we know that $k_B T_I = \frac{GM\overline{m}}{3R_\odot}$: inserting this we find

$$L = 4\pi R_\odot^2 \sigma \left(\frac{GM\overline{m}}{3Rk_B} \right)^4 \frac{\ell}{R} = \frac{(4\pi)^2}{3^5} \frac{\sigma}{k_B^4} G^4 \overline{m}^4 \overline{\rho} \ell M^3 \,. \tag{4.153}$$

The most important part of this result, which is observationally verifiable, is $L \propto M^3$. This allows us to estimate the lifespan of a star: the fraction of mass available to be turned into energy through fusion is a constant multiple of M (slighly less than 1 %): therefore, we have $\tau \propto M/L \propto M^{-2}$. More massive stars die earlier.

The relation $L \propto M^\alpha$ for $\alpha = 3$ is quite close to the data globally, in specific regions we can have better estimates for the powerlaw index: the value $\alpha = 3.5$ is more commonly used for Main Sequence stars (which we will discuss later).

The lifetime always scales like $\tau \propto M^{1-\alpha}$, and we always have $\alpha > 1$, so the fact that more massive stars die earlier always holds.

We then expect that each galaxy there can have been several "generations" of heavy stars, while the very lightest are still in their first generation. This can be observationally confirmed by looking at the abundances of elements, and comparing them with the expected production inside heavy stars.

4.3.2 Thermonuclear fusion

The reaction chain which produces Helium in stellar cores is the following:

$$\mathrm{p} + \mathrm{p} \to \mathrm{d} + \mathrm{e}^+ + \nu_e \tag{4.154a}$$

$$\mathrm{p} + \mathrm{d} \to {}^3\mathrm{He} + \gamma \tag{4.154b}$$

$${}^3\mathrm{He} + {}^3\mathrm{He} \to {}^4\mathrm{He} + \mathrm{p} + \mathrm{p}\,, \tag{4.154c}$$

which involves the weak interaction (for the first process, whose lifetime is $\tau \sim 5 \times 10^9\,\mathrm{yr}$), EM interaction (for the second process: $\tau \sim 1\,\mathrm{s}$) and strong interaction (for the third process: $\tau \sim 3 \times 10^5\,\mathrm{yr}$).

The net balance is 4 protons in, one ${}^4\mathrm{He}$ and some photons and neutrinos out, accounting for the $\approx 0.66\,\%$ mass difference.

We can then see that the weak-interaction part of the chain is the bottleneck.

The chain we had during primordial nucleosynthesis, on the other hand, did not need any weak-interaction processes:

$$\mathrm{n} + \mathrm{p} \to \mathrm{d} + \gamma \tag{4.155a}$$

$$\mathrm{d} + \mathrm{d} \to {}^3\mathrm{He} + \mathrm{n} \tag{4.155b}$$

$${}^3\mathrm{He} + \mathrm{d} \to {}^4\mathrm{He} + \mathrm{p}\,. \tag{4.155c}$$

The issue is that all the free neutrons quickly decayed after primordial nucleosynthesis, and free deuterium is scarce: secondary production chains can then prevail.

The weak interaction chain has a very low power density: $P_\odot / V_\odot \approx 0.27\,\mathrm{W/m^3}$, a lower power density than a human (who generally has a volume less than a cubic meter, but can still output a few hundreds of Watts).

For each ${}^4\mathrm{He}$ nucleus we get an energy $E = (4m_p - m_{{}^4\mathrm{He}})c^2 \approx 25\,\mathrm{MeV}$. Then, we can calculate the number of protons per second the Sun uses in order to produce the power it does.

The rate r of proton usage is given by

$$r = \frac{4L_\odot}{E} \approx 4 \times 10^{38}\,\mathrm{Hz}\,, \tag{4.156}$$

which corresponds to about 6.5×10^{11} kg/s.[6]

For each process one electron neutrino is also emitted, and the first two steps of the process happen twice for each ^{4}He nucleus, so around 2×10^{38} neutrinos per second are produced. These travel basically undisturbed through the Sun and out.

A proton's mass is around $m_p \approx 1\,\mathrm{GeV} \approx 1.78 \times 10^{-27}$ kg, so in the Sun there are around 10^{56} protons: this means that the total lifetime of the Sun will be around 10^{10} years. The Sun is approximately half-way through this lifetime.

4.3.3 Stellar evolution

This section is massively simplified, stellar evolution is complicated, not completely understood and there can be many confounding variables. We will only try to give a general overview.

Throughout the evolution of the star, the interior is near equilibrium between gravitation and the pressure gradient due to the energy emitted by fusion. If the fusion starts to produce more energy, then the star expands and reaches a new equilibrium.

The series of processes which can happen in stellar fusion is shown in table .

Process	Fuel	Products	T_{min}	M_{min}
Hydrogen burning	Hydrogen	Helium	10^7 K	$0.08 M_\odot$
Helium burning	Helium	Carbon, Oxygen	10^8 K	$0.5 M_\odot$
Carbon burning	Carbon	Oxygen, Neon, Sodium	5×10^8 K	$8 M_\odot$
Neon burning	Neon	Magnesium, Oxygen	10^9 K	$9 M_\odot$
Oxygen burning	Oxygen	Magnesium to Sulphur	2×10^9 K	$10 M_\odot$
Silicon burning	Silicon	Iron and nearby elements	3×10^9 K	$11 M_\odot$

Figure 4.1: Solar processes.

As one type of fusion fuel starts to run out, the pressure from the inside diminishes, therefore the interior of the star starts contracting, which typically allows the interior to start fusing the next kind of fuel, while the exterior keeps expanding.

The required temperature to fuse every next element is ever higher, after every cycle there is the possibility of reaching the maximum density allowed by electron degeneracy pressure, which yields a mass threshold for each burning stage, denoted as M_{min} in the table.

For example, the Sun will reach the Helium burning phase, but it will go no further.

A star with a mass of less than $11 M_\odot$ will not reach the iron stage; as it starts burning helium it becomes a red giant, and it only keeps growing as it goes through the burning

[6] Which, to use a gruesome comparison, corresponds to roughly the mass of the entire human population every second.

phases. As it runs out of its last fuel, it starts shrinking and becomes either a white dwarf or a neutron star.

There is a boundary, the Chandrasekhar mass $M_C \approx 1.4 M_\odot$, between the final fate of the star being a white dwarf or a neutron star: this is the maximum mass a star made of protons and electron can have if it must resist its own gravitational collapse only through electron degeneracy pressure.

Now, let us consider massive stars with $M > 11 M_\odot$. They are able to burn oxygen into iron in their cores; now, the thing to note here is that the iron nuclide ^{56}Fe has one of the highest binding energy per nucleon,[7] so once it is reached any successive burning stages would absorb energy, instead of releasing it.

This region in the core will not have any way to provide pressure to counteract the gravity of all the rest of the star above it. When the iron core reaches the Chandrasekhar mass, it collapses onto itself, and the rebound from this creates a supernova, which will either leave a neutron star or a black hole as a remnant.

In the supernova the conditions allow for the formation of heavier elements, beyond iron. The matter which is expelled can form a *planetary nebula*.

Maybe add more details here? Not sure whether it makes sense, maybe refer to the advanced astrophysics notes.

4.3.4 The Hertzsprung-Russel diagram

In the Herzsprung Russel diagram we plot $L/L_\odot$ versus T_{eff}, the latter increasing right to left.

The Main Sequence runs from the upper left to the lower right, we have Red Giants on the upper right and White Dwarves on the lower left. Most of the stars are on the Main Sequence: the hydrogen burning phase lasts a long time.

Pacciani here has a more in-depth discussion of HR diagrams, absolute magnitudes and so on: this would be a useful thing to insert. Also, a figure would be useful.

4.3.5 The interior of a Main Sequence star

We wish to describe the statics of the fluid which makes up a Main Sequence star; the goal we set is to calculate what is the maximum mass of a Main Sequence star — the Main Sequence includes all the stars which are in the process of burning hydrogen.

The density profile The first thing we will need is to calculate the density profile of the star, which is written as $\rho(r)$ since as always we are assuming spherical symmetry.

[7] It is not the highest, since there are a couple nuclides like ^{62}Ni which slightly exceed its binding energy per nucleon. These are collectively known as the "iron group", and are also formed, albeit in smaller amounts, during stellar fusion. For more details, see section IV in Fewell [Few95], a very clear (and not so technical) paper.

Our equation of hydrostatic equilibrium can be written as:

$$\frac{\mathrm{d}P}{\mathrm{d}r} = -\frac{Gm(r)\rho(r)}{r^2} \tag{4.157}$$

$$\frac{r^2}{\rho(r)}\frac{\mathrm{d}P}{\mathrm{d}r} = -Gm(r)\,, \tag{4.158}$$

and we can relate the differential mass of a spherical shell with the differential radius through $\mathrm{d}m = 4\pi r^2 \rho(r)\,\mathrm{d}r$.

Differentiating the equation of hydrostatic equilibrium we find

$$\frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{r^2}{\rho(r)}\frac{\mathrm{d}P}{\mathrm{d}r}\right) = -G\frac{\mathrm{d}m}{\mathrm{d}r} = -4\pi G\rho(r)r^2\,, \tag{4.159}$$

more commonly stated as

$$\frac{1}{r^2}\frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{r^2}{\rho(r)}\frac{\mathrm{d}P}{\mathrm{d}r}\right) = -4\pi G\rho(r)\,. \tag{4.160}$$

This resembles the Laplacian in spherical coordinates, $\nabla^2 f(r) = r^{-2}\partial_r\left(r^2\partial_r f\right)$... can this be stated more precisely?

We can solve this using an equation of state which gives us $P = P(\rho)$; for stellar interiors the cosmological equations of state $P \propto \rho$ do not in general work well, and instead we must generalize to a *polytropic* equation:

$$P = k\rho^{\frac{n+1}{n}} = k\rho^\gamma\,, \tag{4.161}$$

where k and n are constants; also $n = 1/(\gamma - 1)$. This γ is the adiabatic index: for a monoatomic nonrelativistic gas $\gamma = 5/3$ and $n = 3/2$, for an ultrarelativistic gas $\gamma = 4/3$ and $n = 3$.

This will yield a second order differential equation for ρ, which we must complement with two boundary conditions. We can set the value of the central density, $\rho(r = 0) = \rho_c$; also, in order for the density to be a differentiable function of position inside the stare we must also have $\partial_r\rho(r = 0) = 0$: this is because if we move in a straight line through the center of the star, as we pass $r = 0$ we move from a certain value of the derivative to minus that value, since r goes from decreasing to increasing. The only way for this to be continuous is if the derivative is zero.

This is confirmed by the fact that, if we substitute the polytropic equation of state, we find

$$\rho^{1/n}\frac{\mathrm{d}\rho}{\mathrm{d}r} \propto -\frac{Gm(r)}{r^2}\rho(r)\,, \tag{4.162}$$

and the mass in a small region around the origin is approximately $m(r) \sim \rho_c r^3$: therefore $\rho^{1/n}\,\mathrm{d}\rho/\mathrm{d}r \propto \rho_c r$.

These equations can be solved numerically. The radius of the star can be calculated as the one at which the density goes to 0: $\rho(R) = 0$, and the mass of the star is given by $m(R) = M$.

This model is unphysical in that it assumes that the star's interior can be described by a constant γ throughout; as we have seen there are several orders of magnitude of difference in pressure, temperature and density from the core to the surface, so it is a strong assumption to say that it behaves in the same way.

The Clayton model Let us discuss a model proposed by Clayton in 1986, which uses an ansatz for the density profile in order to extract information about the star.

Near the center of the star, we can estimate the mass contained within a spherical shell by assuming $\rho = \rho_c$ throughout, so $m(r) \sim \frac{4\pi}{3}\rho_c r^3$, which we can substitute into the equation of hydrostatic equilibrium:

$$\frac{\mathrm{d}P}{\mathrm{d}r} = -\frac{Gm\rho}{r^2} \approx -\frac{4\pi G}{3}\rho_c^2 r\,, \tag{4.163}$$

so the pressure gradient goes to zero linearly in r.

Also, as r approaches the radius of the star, R, we get $\mathrm{d}P/\mathrm{d}r \to 0$ as well, since the pressure gradient is proportional to $\rho(r)$ in that region, while r approaches a constant.

So, the pressure gradient approaches zero both at the core and at the surface, while in the interior it has a negative value.

The ansatz by Clayton is a relatively simple expression which achieves these requirements:

$$\frac{\mathrm{d}P}{\mathrm{d}r} = -\frac{4\pi}{3}G\rho_c^2 r \exp\left(-\frac{r^2}{a^2}\right), \tag{4.164}$$

where the parameter a has the dimensions of a length, and we take it to be $a \ll R$. This model is quite accurate near the center, not so much near the surface!

By integrating we can calculate the pressure profile:

$$P(r) = \frac{2\pi}{3}G\rho_c^2 a^2\left(\exp\left(-\frac{r^2}{a^2}\right) - \exp\left(-\frac{R^2}{a^2}\right)\right), \tag{4.165}$$

so that the pressure is exactly zero at the surface: $P(R) = 0$.

Substituting the relation $\mathrm{d}m = 4\pi r^2\rho\,\mathrm{d}r$ into the hydrostatic equilibrium equation we find

$$Gm(r)\,\mathrm{d}m = -4\pi r^4\,\mathrm{d}P\,, \tag{4.166}$$

which can be integrated in order to calculate the mass from the pressure profile:

$$\frac{1}{2}Gm^2(r) = -4\pi\int_0^r \mathrm{d}\widetilde{r}\,\widetilde{r}^4\frac{\mathrm{d}P}{\mathrm{d}\widetilde{r}} \tag{4.167}$$

$$m^2(r) = -\frac{8\pi}{G}\left(-\frac{4\pi}{3}\right)G\rho_c^2\int_0^r \mathrm{d}\tilde{r}\,\tilde{r}^5\exp\left(-\frac{r^2}{a^2}\right) \tag{4.168}$$

$$m(r) = \frac{4\pi a^3}{3}\rho_c\Phi(x)\,, \tag{4.169}$$

where we performed a change of variable to $x = r/a$ (bringing out a^6) and defined $\Phi(x)$ as the square root of the dimensionless integral:

$$\Phi^2(x) = 6\int_0^x \mathrm{d}y\,y^5 e^{-y^2} = 6 - 3\Big(x^4 + 2x^2 + 2\Big)e^{-x^2}\,. \tag{4.170}$$

Near the surface x is very large (since, as we mentioned, $a \ll R$), so the exponential dominates: we find $\Phi^2(x) \approx 6$ there.

The density profile can also be expressed in terms of $\Phi(x)$:

$$\begin{aligned}\rho(r) &= \frac{1}{4\pi r^2}\frac{\mathrm{d}m}{\mathrm{d}r} = \frac{1}{4\pi a^2 x^2}\frac{1}{a}\frac{4\pi a^3\rho_c}{3}\frac{\mathrm{d}\Phi}{\mathrm{d}x} &(4.171)\\ &= \frac{\rho_c}{3x^2}\frac{1}{2\Phi}\frac{\mathrm{d}}{\mathrm{d}x}\left(6\int_0^x y^5\exp\Big(-y^2\Big)\,\mathrm{d}y\right) &(4.172)\\ &= \rho_c\left(\frac{x^3 e^{-x^2}}{\Phi(x)}\right), &(4.173)\end{aligned}$$

which, under the assumptions of the model, is a complete description of the density profile. We can also recover the temperature profile from the ideal gas law (which holds as long as the gas inside the star is nonrelativistic):

$$T(r) = \frac{\overline{m}}{k_B}\frac{P(r)}{\rho(r)}\,. \tag{4.174}$$

Let us calculate this near the center of the star, meaning $x \ll 1$;

$$\begin{aligned}\Phi(x) &= \left[6 - 3(x^4 + 2x^2 + 2)\left(1 - x^2 + \frac{x^4}{2} - \frac{x^6}{6}\right)\right] &(4.175)\\ &\approx \left(x^6 - \frac{3}{4}x^8 + \frac{3}{10}x^{10} - \frac{1}{12}x^{12} + \dots\right)^{1/2}, &(4.176)\end{aligned}$$

which we can insert into our expression for $\rho(r)$, also let us expand $P(r)$ to second order, so that we can also calculate $T(r)$:

$$\rho(r) \approx \rho_c\left(1 - \frac{5}{8}\frac{r^2}{a^2} + \dots\right) \tag{4.177}$$

$$P(r) \approx \frac{2\pi}{3}G\rho_c^2 a^2\left(1 - \frac{r^2}{a^2} + \dots\right) \tag{4.178}$$

$$T(r) \approx T_c\left(1 - \frac{3}{8}\frac{r^2}{a^2} + \dots\right), \tag{4.179}$$

where

$$T_c = \frac{\overline{m}}{k_B}\frac{2\pi}{3}G\rho_c a^2 . \tag{4.180}$$

Moving to the surface, we can calculate the total mass

$$M = m(R) = \frac{4\pi\rho_c a^3}{3}\Phi(R/a) \approx \frac{4\pi\rho_c a^3\sqrt{6}}{3} . \tag{4.181}$$

Then, the average density is given by

$$\langle\rho\rangle = \frac{M}{\frac{4\pi}{3}R^3} = \sqrt{6}\left(\frac{a}{R}\right)^3\rho_c , \tag{4.182}$$

so if it is the case that $a \ll R$ then we also have $\rho_c \gg \langle\rho\rangle$.

We can invert this relation to find a in terms of M and ρ_c; for the Sun we find $a \approx R_\odot/5.4$, $\rho(a) = 0.53\rho_c$ and $m(a) = 0.28M_\odot$.

This means that, as we expected, the Sun is quite concentrated: over a quarter of its mass is contained within $(1/5.4)^3 \approx 0.6\,\%$ of its volume.

A useful result we can derive from this model is the central pressure expressed in terms of the mass and central density:

$$P_c = \frac{2\pi}{3}G\rho_c^2 a^2 = \frac{2\pi}{3}G\rho_c^2\left(\frac{3M}{4\pi\rho_c\sqrt{6}}\right)^{2/3} \tag{4.183}$$

$$\approx \left(\frac{\pi}{36}\right)^{1/3} GM^{2/3}\rho_c^{4/3} . \tag{4.184}$$

This model then predicts the prefactor $q = (\pi/36)^{1/3} \approx 0.44$. The powers of M and ρ_c are the same in more sophisticated models.

From simulations with varying γ we get: for $\gamma = 5/3$ the factor is $q \approx 0.48$, for $\gamma = 4/3$ the factor is $q \approx 0.36$. The results are quite close to ours!

Pacciani also states that $q < (\pi/6)^{1/3} \approx 0.14\ldots$ not sure how that would make sense!

4.3.6 The maximum mass

We can apply the ideal gas relation to the core of the star: then we find

$$k_B T_c = \overline{m}\frac{P_c}{\rho_c} \approx \left(\frac{\pi}{36}\right)^{1/3} G\overline{m}M^{1/3}\rho_c^{1/3} . \tag{4.185}$$

In the core of the star, we have both nonrelativistic and relativistic material in equilibrium: electrons and protons are nonrelativistic, while photons are relativistic. We have discussed earlier that if most of the material in a star were relativistic it would become unstable (since, as $\gamma \to 4/3$, the binding energy approaches 0); let us then discuss the composition of the star.

We can decompose the pressure at the core into the fractions due to nonrelativistic matter and to radiation:

$$P_c = P_m + P_r = \beta P_c + (1-\beta) P_c \,, \tag{4.186}$$

where we define $\beta \in (0,1)$ as the fraction of the core pressure which is due to matter: $\beta = P_m / P_c$.

The two contributions can be separately expressed as:

$$\beta P_c = P_m = \frac{\rho_c k_B T_c}{\overline{m}} \tag{4.187}$$

$$(1-\beta) P_c = P_r = \frac{1}{3} a T^4 \,, \tag{4.188}$$

where a is the radiation constant, related to the Stefan-Boltzmann constant σ:

$$a = \frac{\pi^2 k_B^2}{15 \hbar^3 c^3} \,. \tag{4.189}$$

We can then relate β to the mass of the star: in order to simplify the core temperature, we start by computing

$$\frac{(\beta P_c)^4}{(1-\beta) P_c} = \frac{\rho_c^4}{\overline{m}^4} (k_B T_c)^4 \frac{3}{a T_c^4} \tag{4.190}$$

$$\frac{\beta^4}{1-\beta} P_c^3 = \frac{3}{a} \left(\frac{k_B \rho_c}{\overline{m}} \right)^4 \,, \tag{4.191}$$

which we can invert to find an expression for the core pressure P_c in terms of β, which we then compare to the expression we found for the core pressure as a result of the Clayton model:

$$P_c = \left(\frac{3}{a} \frac{1-\beta}{\beta^4} \right)^{1/3} \left(\frac{k_B \rho_c}{\overline{m}} \right)^{4/3} = \left(\frac{\pi}{36} \right)^{1/3} G M^{1/3} \rho_c^{4/3} \tag{4.192}$$

$$\left(\frac{\pi}{36} \right)^{1/3} G M^{2/3} = \left(\frac{3}{a} \frac{(1-\beta)}{\beta^4} \right)^{1/3} \left(\frac{k_B}{\overline{m}} \right)^{4/3} \,, \tag{4.193}$$

the core density simplifies!

So, if we compare stars at the same stage of fusion so that $\overline{m}$ is constant, we have $M \propto f(\beta) = (1-\beta)^{1/2} / \beta^2$.

$f(\beta)$ decreases as β increases, and it diverges to $+\infty$ for $\beta \to 0$.

Looking at the plot the other way, the heavier the star, the larger the contribution of radiation to the core pressure, which is what $1-\beta$ quantifies.

Figure made quickly, to improve by scaling it correctly, making it vector, setting the text in the right font.

This makes sense intuitively: heavier stars reach higher temperatures and densities, so they have more radiation in the core.

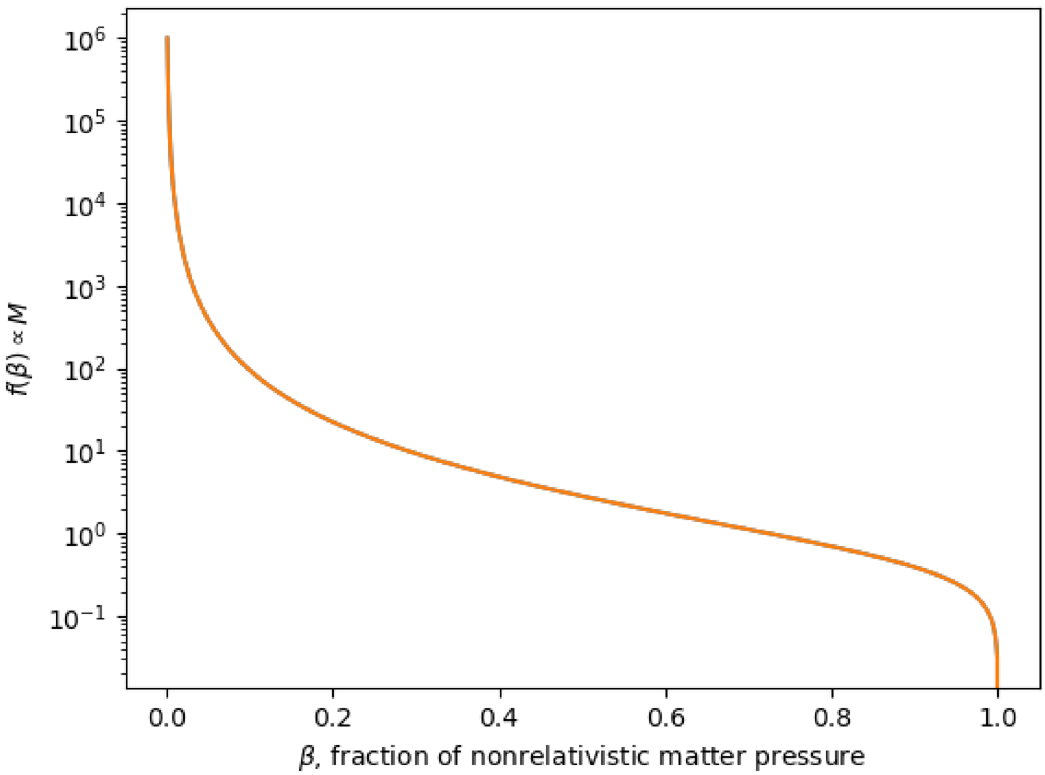

Figure 4.2: A plot of $f(\beta)$.

We know that for $\beta \to 0$ the star is surely unstable, but the instability is actually reached earlier, since even before the gravitational binding energy being exactly zero large parts of the star can be flung out as stellar winds. Proper considerations about what an appropriate critical value of β should be allow us to bound the stellar mass from above, at around $50 M_\odot$.

4.3.7 Degenerate electron gas

Now, we will deal with the degenerate electron gas in stars, and see what is its effect on the minimum and maximum mass of a star.

The distribution function of the electrons, which are fermions, is given by

$$f(p) = \left[\exp\left(\frac{\epsilon_p - \mu}{k_B T} \right) + 1 \right]^{-1}, \tag{4.194}$$

where $\epsilon_p = \sqrt{m^2 c^4 + p^2 c^2}$. With it, we can calculate the number density of electrons:

$$n_e = \frac{g_s}{h^3} \int \mathrm{d}^3 p \, f(p) \,, \tag{4.195}$$

where g_s, the number of helicity states of the electron, is equal to 2.

We want to consider the degenerate case for this distribution, which corresponds to the saturation of all the low-energy configurations in phase space: this is known as a Fermi gas. As the temperature approaches zero, the phase space distribution approaches the configuration

$$f(p) = \lim_{T\to 0}\left[\exp\left(\frac{\epsilon_p - \mu}{k_B T}\right) + 1\right]^{-1} = \begin{cases} 1 & \epsilon_p < \mu \\ 0 & \epsilon_p > \mu\,. \end{cases} \tag{4.196}$$

The chemical potential yields a critical energy, known as the Fermi energy, $\epsilon_F = \mu$, which is also tied to a Fermi momentum:

$$\epsilon_F^2 = c^2 p_F^2 + m^2 c^4\,. \tag{4.197}$$

The number density of electrons in this configuration is given by the integral mentioned before: since the distribution is spherically symmetric we have

$$n_e = 2\int_0^{p_F} \mathrm{d}p\, p^2 4\pi \frac{1}{h^3} = \frac{8\pi}{3}\left(\frac{p_F}{h}\right)^3, \tag{4.198}$$

which allows us to express the Fermi momentum in terms of the number density of electrons:

$$p_F = \left(\frac{3n_e}{8\pi}\right)^{1/3} h\,. \tag{4.199}$$

In natural units, this is roughly $p_F \approx 6.6\sqrt[3]{n_e}$.

The energy density is given by the expression

$$\rho = \frac{2}{h^3}\int_0^{p_F} 4\pi p^2\,\mathrm{d}p\,\epsilon_p\,, \tag{4.200}$$

which we can consider in either the nonrelativistic or the ultrarelativistic limit — the analytic integral is complicated and not very enlightening.

Nonrelativistic limit In this limit the energy is approximately

$$\epsilon_p = mc^2 + \frac{p^2}{2m}\,, \tag{4.201}$$

so in the computation we need to integrate a polynomial: the result is

$$\rho = n\left(mc^2 + \frac{3}{10}\frac{p_F^2}{m}\right), \tag{4.202}$$

where the first term corresponds to the rest-energy of the electrons, while the second gives their kinetic energy. We have derived earlier the following expression for the pressure of a nonrelativistic gas:

$$P = \frac{2}{3}\frac{E_k}{V}\,, \tag{4.203}$$

and E_k/V is precisely the kinetic energy density, the second term of the expression for the total energy density ρ; so for our nonrelativistic Fermi gas we will have:

$$P = n\frac{p_F^2}{5m}\,. \tag{4.204}$$

Since the Fermi momentum p_F can be written as a function of the number density n_e, so can the pressure P: we find

$$P = \underbrace{\frac{h^2}{5m}\left(\frac{3}{8\pi}\right)^{2/3}}_{K_{NR}} n_e^{5/3}\,. \tag{4.205}$$

Ultra relativistic limit In the relativistic case, on the other hand, we can approximate the energy as $\epsilon_p \approx cp$, so the energy density will be given by

$$\rho = \frac{3}{4}n\rho_F c\,. \tag{4.206}$$

In this case, we also know that the pressure becomes

$$P = \frac{1}{3}\frac{E_k}{V}\,, \tag{4.207}$$

so we find

$$P = \underbrace{\frac{hc}{4}\left(\frac{3}{8\pi}\right)^{1/3}}_{K_{UR}} n^{4/3}\,. \tag{4.208}$$

Fermion gas classification We have discussed some expressions describing a non-relativistic or ultrarelativistic degenerate fermion gas.

We have derived our results with the assumption $T \to 0$, but a gas can behave very similarly with nonzero temperatures as well. What is the temperature threshold under which the gas behaves in a degenerate-like way? We will not discuss how the transition region looks, but if $k_B T \ll \epsilon_F$ then the gas behaves like a degenerate one, while if $k_B T \gg \epsilon_F$ then there will be many unfilled gaps in the phase space distribution, so the gas will not be degenerate.

Recall that $p_F \propto n_e^{1/3}$: therefore, in a log-log plot of temperature T versus density of possible electron densities n_e we can draw a line distinguishing the degenerate and nondegenerate cases, with the critical temperature becoming higher for higher n_e.

Is it really a straight line though? If the criterion is indeed to compare $k_B T$ and ϵ_p then I'd expect a curve, since ϵ_F is not a polynomial function of p_F...

Having distinguished the degenerate and nondegenerate regions, we can distinguish the relativistic and nonrelativistic ones: for the nondegenerate case, as is usual, we reach the relativistic condition if we increase the temperature.

In the degenerate case this is not really the case: as long as the gas is degenerate, the temperature does not really matter, and the gas becomes relativistic when the Fermi energy ϵ_F becomes larger than the mass of the fermion. Since ϵ_F is only a function of the number density, this means that the gas can become relativistic at arbitrarily low temperatures as long as it is dense enough.

Add plot — maybe we can do a chromatic region plot, integrating numerically the distribution and coloring the region based on the fraction of relativistic particles, and for the degeneracy measure in some way how "sharp" the boundary is between the filled and unfilled regions?

Application to the Sun As we have discussed earlier, the core temperature, pressure and density of the Sun are related by the following relation:

$$P_c = \frac{\rho_c}{\overline{m}} k_B T_c \,. \tag{4.209}$$

The average mass which appears here is a function of the chemical composition of the interior: we can neglect all the metals and only consider the mass fractions of hydrogen (x_1) and of helium (x_4):

$$\overline{m} = 2 m_H \times \frac{1}{1 + 3x_1 + 0.5 x_4} \,. \tag{4.210}$$

This expression works well in the limits of $x_1 = 1$ and $x_4 = 1$, but where does it come from? I would have expected $\overline{m} = (x_1 m_H + x_4 m_{He})/2 \dots$

The Clayton model gave us an expression for the central pressure P_c, which we turned into one for the central temperature T_c:

$$P_c = \approx \left(\frac{\pi}{36}\right)^{1/3} G M^{2/3} \rho_c^{4/3} \tag{4.211}$$

$$k_B T_c \approx \left(\frac{\pi}{36}\right)^{1/3} G \overline{m} M^{2/3} \rho_c^{1/3} \,, \tag{4.212}$$

however when deriving it we not consider the effect of the fact that the gas there may be at least party degenerate.

Let us consider a different approximation: suppose that the electrons in the core are fully degenerate and nonrelativistic, while the ions (whose density is n_i, which by local neutrality is also equal to $n_e = \rho_c / \overline{m}$) are completely classical.[8] Our estimate for the central pressure

[8] In order to see why it makes sense to consider them as classical while the electrons are degenerate, let us look at the fact that in the nonrelativistic approximation the Fermi energy is given by $\epsilon_F = p_F^2/2m \sim m^{-1} n^{2/3}$, so the critical temperature needed for a Fermi gas to become degenerate depends on the number density as well as the mass of the particle: for a higher-mass particle, the Fermi energy is lower.

The electrons being degenerate means that the temperature of the core is (roughly speaking) lower than their Fermi energy; the Fermi energy of the ions however is at least three orders of magnitude lower, so it makes sense that it is not as low as the Fermi temperature of the ions.

will need to account for both electrons and ions:

$$P_c = k_{NR} n_e^{5/3} + n_i k_B T_c \,. \tag{4.213}$$

Let us equate this expression with the one given by the Clayton model for the central pressure: we find

$$\left(\frac{\pi}{36}\right)^{1/3} G M^{2/3} \rho_c^{4/3} = k_{NR} \left(\frac{\rho_c}{m_H}\right)^{5/3} + \frac{\rho_c}{m_H} k_B T_c \tag{4.214}$$

$$k_B T_c = \underbrace{\left(\frac{\pi}{36}\right)^{1/3} G m_H M^{2/3}}_{A} \rho_c^{1/3} - \underbrace{k_{NR} m_H^{-2/3}}_{B} \rho_c^{2/3} \,. \tag{4.215}$$

We can then ask what is the maximum temperature T_c we can reach for a given mass M if we vary the core density ρ_c: this can be calculated to be

$$\rho_c^{\max} = (A/2B)^3 \approx 5 \times 10^7 \,\mathrm{kg/m^3} \left(\frac{M}{M_\odot}\right)^2 , \tag{4.216}$$

where we have

$$k_B T_c = \frac{A^2}{4B} = \left(\frac{\pi}{36}\right)^{2/3} \frac{G^2 m_H^{8/3}}{4 k_{NR}} M^{4/3} \approx 5.7 \,\mathrm{keV} \left(\frac{M}{M_\odot}\right)^{4/3} . \tag{4.217}$$

This allows us to estimate the minimum mass a star needs to have in order to fuse hydrogen: we just need to set T_c to be equal to the ignition temperature $T_c = T_{\mathrm{ign}} \approx 1\,\mathrm{keV}$ and we find

$$M_{\min} = \left(\frac{36}{\pi}\right)^{1/2} \left(\frac{4 k_{NR}}{G^2 m_H^{8/3}}\right)^{3/4} \left(k_B T_{\mathrm{ign}}\right)^{3/4} \approx 0.27 \,\mathrm{M}_\odot \,. \tag{4.218}$$

This is a much better estimate than the one we found earlier since we are now accounting for the degenerate Fermi gas nature of the electrons in the core (this lowers the estimate, since it means that even at relatively low temperatures there will be electrons with high energy) and since we are computing the core density ρ_c instead of the average density $\overline{\rho}$.

The estimate is... still not great really, right? it is still 3 times larger than the correct value of $0.08 M_\odot$! How do we account for such a discrepancy?

In order to have some numbers at hand, with a number density like that of the core of the Sun the Fermi temperature for electrons is $\sim 11\,\mathrm{MK}$, while the Fermi temperature for protons is a measly $\sim 6000\,\mathrm{K}$. The actual temperature of the core is $T_c \sim 15\,\mathrm{MK}$, slightly above the Fermi temperature of electrons. It is close enough that modelling them as degenerate works, while the assumption of the ions being nondegenerate is completely valid.

Expressing the result with coupling constants The gravitational potential energy between two hydrogen nuclei separated by a distance equal to their (reduced!) Compton wavelength $r = \hbar / m_H c$ is

$$E_g = -\frac{G m_H^2}{r} = -\frac{G m_H^3 c}{\hbar} . \tag{4.219}$$

Comparing this to the rest energy of an electron, $E = m_H c^2$, is the way to calculate the *gravitational coupling constant* α_G, a dimensionless parameter quantifying the "strength" of the gravitational interaction between hydrogen nuclei:

$$\alpha_G = \frac{E_g}{E} = \frac{G m_H^2}{\hbar c} \sim 5.9 \times 10^{-39} . \tag{4.220}$$

In natural units, $\alpha_G = m_H^2 / m_P^2$.

By a similar line of reasoning we find the electromagnetic coupling constant:

$$\alpha_{EM} = \frac{e^2}{4\pi\epsilon_0 \hbar c} \approx \frac{1}{137} , \tag{4.221}$$

which is *enormously* greater.

In terms of the gravitational coupling constant the minimum mass we found can be written as

$$M_{\text{min}} \approx 16 \left(\frac{k_B T_{\text{ign}}}{m_e c^2} \right)^{3/4} \alpha_G^{-3/2} m_H . \tag{4.222}$$

If $T_{\text{ign}} \sim 1.5 \times 10^6\,\text{K}$, one tenth of the temperature of the Sun, we find

Is 0.1 keV really enough to reach ignition? This seems to contradict what was said earlier...

$$M_{\text{min}} \sim 0.03 \alpha_G^{-3/2} m_H . \tag{4.223}$$

We can apply a similar line of reasoning to the formula we found for the maximum mass: taking equation (4.193) with a critical fraction of nonrelativistic matter of $\beta = 0.5$ and $\overline{m} = 0.61 m_H$ we get a result which, once again, scales with $\alpha_G^{-3/2} m_H$:

$$M_{\text{max}} \approx 56 \alpha_G^{-3/2} m_H . \tag{4.224}$$

This hints to the fact that $m_* = \alpha_G^{-3/2} m_H$ is an important characteristic mass for all of stellar evolution.

This is around $1.85 M_\odot$, and it corresponds to a number of nucleons of

$$N_* = \frac{m_*}{m_H} \approx 2 \times 10^{57} . \tag{4.225}$$

4.4 Stellar remnants

4.4.1 Full degeneracy and white dwarfs

White dwarfs are the remnants of low-mass stars who have exhausted the elements they are able to fuse in their core. They glow, emitting thermal radiation, which causes their temperature to slowly decrease until they become brown dwarfs. They are dim but observable; the closest one to the Solar System is Sirius B, a companion to the brightest star in the night sky.

They are of interest to us since they allow us to apply the theory of degenerate Fermi gasses once more: they are very dense objects, since there is no fusion-induced pressure gradient inside them to balance gravity, and the electrons inside them form a degenerate gas.

The number density of electrons inside a white dwarf is given by

$$n_e = Y_e \frac{\rho_c}{m_H}\,, \tag{4.226}$$

where $Y_e = (1 + x_1)/2$ quantifies the number of electrons per baryon (x_1 is the hydrogen mass fraction).

Why would Y_e be given by that expression? hydrogen has one electron per each baryon, but helium also has half an electron per baryon... If the white dwarf was exclusively hydrogen, would we not expect $Y_e = 1/2$?

Let us start by assuming that the matter is nonrelativistic: then the pressure is given by

$$P = k_{NR} n_e^{5/3} = k_{NR} \left(\frac{Y_e \rho_c}{m_H} \right)^{5/3}\,, \tag{4.227}$$

which as usual we compare to the results of the Clayton model:

$$P_c = \left(\frac{\pi}{36} \right)^{1/3} G M^{2/3} \rho_c^{4/3}\,. \tag{4.228}$$

Equating these two we find

$$\rho_c \approx \frac{3.1}{Y_e^5} \left(\frac{M}{m_*} \right)^2 \frac{m_H}{(h/m_e c^2)^3}\,. \tag{4.229}$$

If, on the other hand, we were to assume that the matter is ultrarelativistic the pressure would be given by

$$P = k_{UR} n_e^{4/3} = k_{UR} \left(\frac{Y_e \rho_c}{m_H} \right)^{4/3}\,, \tag{4.230}$$

so, instead of getting an expression for the central density ρ_c, we would find

$$k_{UR} \left(\frac{Y_e \rho_c}{m_H} \right)^{4/3} \approx \left(\frac{\pi}{36} \right)^{1/3} G M^{2/3} \rho_c^{4/3} : \tag{4.231}$$

in this limit the expression becomes independent of ρ_c!

This gives us a limit mass, since as we increase the mass of a white dwarf which is not relativistic we increase its density and thus its temperature, making it closer to being relativistic, and this is the mass we get for the fully relativistic configuration (which is unstable because of the usual binding energy considerations).

The limit is known as the Chandrasekhar mass, the largest mass at which a fully degenerate white dwarf can support itself:

$$M_{CH} = \left(\frac{36}{\pi}\right)^{1/2} \left(\frac{Y_e}{m_H}\right)^2 \left(\frac{k_{UR}}{G}\right)^{3/2} \approx 2.3 Y_e^2 m_* \approx 4.3 Y_e^2 M_\odot \approx 1.4 M_\odot \,. \tag{4.232}$$

4.4.2 The Chandrasekhar limit in more detail

We want to derive the Chandrasekhar limit in a more precise manner.

Instead of approximating the gas as either ultrarelativistic or nonrelativistic, we can use the correct expression for the particle energy in the integral for the momentum:

$$P = \frac{4\pi}{3h^3} g_* \int_0^{p_F} \mathrm{d}p \, p^2 \frac{p^2 c^2}{\epsilon_p} \,, \tag{4.233}$$

with $\epsilon_p = \left(p^2c^2 + m^2c^4\right)^{1/2}$.

We change variables to the dimensionless $x = p/(m_e c)$ and substitute $g_* = 2$, since electrons have spin 1/2:

$$P = \frac{8\pi}{3h^3} m_e^4 c^5 \int_0^{x_F} \frac{x^4}{(1+x^2)^{1/2}} \, \mathrm{d}x \,. \tag{4.234}$$

The variable x_F is given by:

$$x_F = \frac{p_F}{m_e c} = \left(\frac{3n_e}{8\pi}\right)^{1/3} \frac{h}{m_e c} = \left(\frac{3 Y_e \rho_c}{8\pi m_H}\right)^{1/3} \frac{h}{m_e c} \,, \tag{4.235}$$

and, since the electrons are fully degenerate, $x_F \gg 1$ corresponds to the ultrarelativistic case while $x_F \ll 1$ corresponds to the nonrelativistic case.

Now, as $x_F \to \infty$ the integral is asymptotically

$$\int_0^{x_F} \frac{x^4}{\sqrt{1+x^2}} \, \mathrm{d}x \sim \frac{x_F^4}{4} \,, \tag{4.236}$$

so we define

$$I(x_F) = \frac{4}{x_F^4} \int_0^{x_F} \frac{x^4}{(1+x^2)^{1/2}} \, \mathrm{d}x \tag{4.237}$$

$$= \frac{3}{2x^4} \left(x(1+x^2)^{1/2} \left(\frac{2x^2}{3} - 1 \right) + \log\left(x + (1+x^2)^{1/2} \right) \right) \,, \tag{4.238}$$

which approaches 1 as $x_F \to \infty$. Then, we can write the pressure as

$$P = \frac{8\pi}{3h^3} m_e^4 c^5 \frac{x_F^4}{4} I(x_F) \tag{4.239}$$

$$= k_{UR} n_e^{4/3} I(x_F)\,. \tag{4.240}$$

We can see that this manipulation works by explicitly doing the calculation, but we can also just observe that the limiting case $x_F \to \infty$ must reduce to the ultrarelativistic approximation, so the prefactor must be the same.

In the nonrelativistic case, $x_F \ll 1$, we have $I(x_F) \sim 4x_F/5$, and since $x_F \sim n_e^{1/3}$ this yields $P \sim n_e^{5/3}$ as expected. This expression *interpolates* between the two limits.

Then, we can apply the same reasoning as before: we compare the pressure of the Fermi gas with the prediction of the Clayton model to find

$$k_{\mathrm{UR}} \left(\frac{Y_e \rho_c}{m_H} \right)^{4/3} I(x_F) \approx \left(\frac{\pi}{36} \right)^{1/3} G M^{4/3} \rho_c^{4/3}\,, \tag{4.241}$$

so we can extract the mass:

$$M = I(x_F)^{3/2} M_{\mathrm{Ch}}\,, \tag{4.242}$$

where $M_{\mathrm{Ch}} \approx 1.4\,\mathrm{M}_\odot$ is the Chandrasekhar mass we defined earlier.

Quick and dirty Mathematica plot, to update (and flip the axes maybe).

We can see that as we increase the mass approaching M_{Ch} the central density diverges: this is not physically possible, of course, so as it gets higher some usually-prohibited process takes over; typically for white dwarfs this is electron capture, by which electrons and protons combine into neutrons, forming a neutron star.

We have not used any particular characteristics of electrons beyond their being fermions, so this line of reasoning may be used to also bound the mass of a neutron star, since neutrons are fermions as well. The issue, as we will see, is that general-relativistic corrections become important in that case, since neutron stars have a much higher density.

4.4.3 White dwarf characteristics

Now that we have a model for the equation of state at the core of a fully degenerate object like a white dwarf, we can try to extract some of its characteristics: it is reasonable (from more complete studies of the object) to estimate the mean density as 1/6 of the central one,

$$\langle \rho \rangle = \frac{1}{6} \rho_c = \frac{0.51}{Y_e^2} \left(\frac{M}{m_*} \right)^2 \frac{m_H}{(h/m_e c)^3}\,, \tag{4.243}$$

so we can estimate the radius as

$$R = \left(\frac{3M}{4\pi \langle \rho \rangle} \right)^{1/3} \approx 0.77 Y_e^{5/3} \left(\frac{M}{m_*} \right)^{1/3} \underbrace{\alpha_G^{-1/2} \frac{h}{m_e c}}_{\ell_{WD}}\,, \tag{4.244}$$

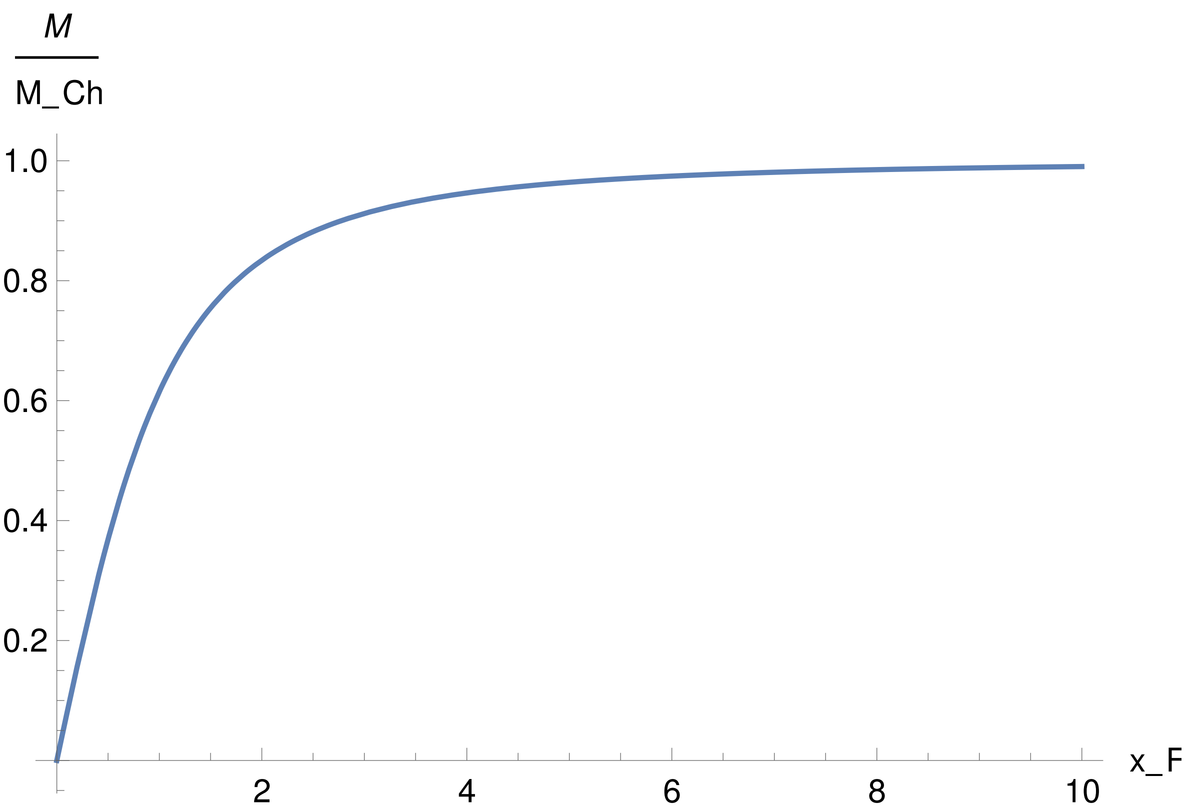

Figure 4.3: A plot of M/M_{Ch} against $x_F \propto n_e^{1/3} \propto \rho_c^{1/3}$.

which is of the same order of magnitude as the characteristic length

$$\ell_{WD} = \alpha_G^{-1/2} \frac{h}{m_e c} \approx 3 \times 10^7 \,\text{m} \approx 0.04 R_\odot \,, \tag{4.245}$$

so, taking $Y_e = 0.5$ we can express the radius as

$$R = \frac{R_\odot}{74} \left(\frac{M_\odot}{M} \right)^{1/3} . \tag{4.246}$$

Using the radius we can estimate the luminosity of the thermal radiation emitted by these bodies:

$$L = 4\pi R^2 \sigma T_E^4 = \frac{1}{74^2} \left(\frac{M_\odot}{M} \right)^{4/3} \left(\frac{T_E}{6000\,\text{K}} \right) L_\odot \,, \tag{4.247}$$

so if we take a typical effective temperature of around 10^4 K (recall that white dwarfs are in the blue part of the HR diagram), $M = 0.4 M_\odot$ we get $L \approx 3 \times 10^{-3} L_\odot$: they are very dim.

4.4.4 Neutron stars

The first thing to consider when discussing neutron stars is the fact that neutrons, as we discussed, are usually unstable, with lifetimes on the order of 10 min. How can a neutron star be stable then? The process through which neutrons decay is:

$$\mathrm{n} \to \mathrm{p} + \mathrm{e}^- + \overline{\nu}_e\,, \tag{4.248}$$

and the crucial fact is that neutron stars are composed of a degenerate neutron gas *as well as* a degenerate *ultrarelativistic* electron gas: the Fermi temperature is much higher than $m_e c^2/k_B$ [Yak+, eq. 1]. While the electron gas is ultrarelativistic the neutron gas is not; the energy released by neutron decay is of the order of 800 keV, so the momentum an emitted electron would have would be well within the Fermi sphere, which is already full!

Thus, neutron decay is inhibited; on the other hand, electron capture, which looks like

$$\mathrm{e}^- + \mathrm{p} \to \mathrm{n} + \nu_e\,, \tag{4.249}$$

is favoured, and it can increase the number of neutrons.

We can look at the Saha formula to get numerical estimates for the equilibrium between these processes. The chemical potential of neutrinos can be neglected, therefore we find

$$\mu_\mathrm{n} = \mu_\mathrm{p} + \mu_\mathrm{e}\,, \tag{4.250}$$

and, since as we saw earlier the chemical potential of a Fermi gas is its Fermi energy, the same equation holds for their Fermi energies:

$$\epsilon_{F,\mathrm{n}} = \epsilon_{F,\mathrm{p}} + \epsilon_{F,\mathrm{e}}\,. \tag{4.251}$$

How does this translate into the number densities of the three constituents? We have the constraint that the number density of protons must equal the number density of electrons in order to ensure local neutrality, while there is no constraint on the ratio between neutrons and protons.

We will not get into the calculation, but due to the slight mass imbalance $m_n > m_p$ we have a large difference in the number densities: typically,

$$n_\mathrm{p} = n_\mathrm{e} = \frac{n_\mathrm{n}}{200}\,. \tag{4.252}$$

Since the overwhelming majority of the particles in the neutron star are neutrons, we can make our calculations with the approximation that the number of neutrons per baryon is $Y_n \approx 1$:

$$n_\mathrm{n} = Y_n \frac{\rho_c}{m_\mathrm{n}} \approx \frac{\rho_c}{m_\mathrm{n}}\,. \tag{4.253}$$

Typical values for these densities are $\rho_c \approx 2 \times 10^{17}\,\mathrm{kg/m^3}$ and $n_\mathrm{n} \approx 10^{44}\,\mathrm{m^{-3}}$.

With a similar reasoning to the one we applied to white dwarfs we can calculate

$$\rho_c^{\mathrm{NS}} \approx 3.1 \left(\frac{M}{M_*}\right)^2 \frac{m_n}{(h/m_n c)^3}\,, \tag{4.254}$$

which, due to the fact that $m_n \gg m_e$, is much larger than the corresponding result for white dwarfs

$$\rho_c^{WD} \approx \frac{3.1}{Y_e^5} \left(\frac{M}{M_*}\right)^2 \frac{m_H}{(h/m_e c^2)^3}\,. \tag{4.255}$$

In both cases, we used the characteristic mass $M_* = \alpha_G^{-3/5} m_n \approx 1.85 M_\odot$.

As we did before, knowing the central density we can estimate the average one, which then allows us to calculate the radius:

$$R = 0.77 \left(\frac{M_*}{M} \right)^{1/3} \alpha_G^{-1/2} \frac{h}{m_n c} , \tag{4.256}$$

where the characteristic length is given by

$$L_n = \alpha_G^{-1/2} \frac{h}{m_n c} \approx 17\,\text{km} \approx \frac{1}{1200} L_e , \tag{4.257}$$

1200 times smaller than the corresponding length scale for white dwarfs (denoted with an e for "electron degeneracy").

Finally, we can compute a maximum mass: $M_{\text{max}}^{\text{NS}} = 3.1 M_* = 5.8 M_\odot$.

We have neglected general-relativistic effects, but would they be relevant? The quantity we need to compute is the ratio of the Schwarzschild radius to the actual radius of the NS:

$$\frac{R_{\text{Schw}}}{R} = \frac{2GM}{Rc^2} \approx 0.4 \left(\frac{M}{M_*} \right)^{4/3} , \tag{4.258}$$

which is large, of order 1! We have not computed a minimum mass for a neutron star, but typically their mass is of the order of the Chandrasekhar mass, $M^{NS} \approx 1.4 M_\odot$ because of how they form in supernovae (so, $M/M_* \sim 1$). The neutron star might not be small enough to actually collapse into a black hole, but surely general relativity must be considered when describing its dynamics.

Neutron stars were first detected as very regular radio pulses: *pulsars*. These are due to the very strong ($\sim 10^8$ T) magnetic fields accelerating particles in beams aligned with the magnetic poles of the NS; these are not aligned with the rotation axis of the NS, so they constantly change the direction of emission, and the Earth can happen to be in this cone.

In order to estimate how fast these pulses can be (in a classical and rough way), let us assume that the NS is rotating barely below a speed which would disintegrate it, so that its binding energy equals its rotational energy:

$$\frac{GM^2}{R^2} \approx R \omega_{\text{max}}^2 , \tag{4.259}$$

we can then compute the minimum period using the expression we have for the radius in terms of M and M_*:

$$\tau_{\text{min}} = \frac{2\pi}{\omega_{\text{max}}} \approx 2\pi \left(\frac{R^3}{GM} \right)^{1/2} \approx 11 \left(\frac{M_*}{M} \right) \alpha_G^{-1/2} \frac{h}{m_n c^2} \approx 0.6 \frac{M_*}{M} \text{ms} . \tag{4.260}$$

The signals produced by pulsars are of this order of magnitude — we have observed "millisecond pulsars", so NSs do indeed rotate close to these extremely high rates.

Neutron stars can also produce gravitational waves in their rotation, as long as they have a slight asymmetry (a "mountain", although their typical sizes are of the order of centimeters); these would have frequencies in the Hz range. We cannot detect these with ground-based detectors, but we might be able to do so with space-based ones.

4.4.5 Relativistic corrections to the equation of state

Black holes are objects which are so dense that their radius is smaller than the Schwarzschild radius:

$$R < \frac{2GM}{c^2} = R_{\text{Sch}}\,. \tag{4.261}$$

General Relativity predicts that when this is the case an **event horizon** form, a surface which forms a causal boundary: as is almost cliché, *not even light can escape*.

Let us see how the classical description of a star fails for objects with relativistic masses. The equation of hydrostatic balance, derived under classical assumptions, reads:

$$\frac{\mathrm{d}P}{\mathrm{d}r} = -\frac{Gm\rho}{r^2}\,. \tag{4.262}$$

If we seek a relativistic analogue under similar assumption (spherical symmetry and equilibrium) we find the **Tolman-Oppenheimer-Volkov** equation:

$$\frac{\mathrm{d}P}{\mathrm{d}r} = -\frac{Gm\rho}{r^2}\left(1+\frac{P}{\rho c^2}\right)\left(1+\frac{4\pi r^3 P}{mc^2}\right)\left(1-\frac{2Gm}{rc^2}\right)^{-1}\,. \tag{4.263}$$

This equation is exact under the assumptions we mentioned. In the classical limit ($2Gm/c^2 \ll r$ and $P \ll \rho c^2$) this reduces to the classical hydrostatic balance equation. The first correction is reminiscent of the second Friedmann equation: *the pressure itself contributes to the inertia of the system*.

Let us see how their predictions differ assuming constant density, $\rho \equiv \rho_0$. In the Newtonian case the mass below a radius r is given by

$$m(r) = \frac{4\pi}{3}\rho_0 r^3\,, \tag{4.264}$$

using which we can integrate the hydrostatic balance equation (from the surface, where the pressure vanishes) to find:

$$P(r) = \int_R^{r_0} \underbrace{\left(-\frac{Gm\rho}{r^2}\right)}_{\mathrm{d}P/\mathrm{d}r} \mathrm{d}r = \frac{2\pi G}{3}\rho_0^2\left(R^2 - r_0^2\right)\,. \tag{4.265}$$

Then, the central pressure is given by

$$P_c^{\text{classical}} = \frac{2\pi}{3}G\rho_0^2 R^2 = \left(\frac{\pi}{6}\right)^{1/3} GM^{2/3}\rho_c^{4/3}\,. \tag{4.266}$$

Keeping the constant-density assumption, this can be done analytically in the relativistic case as well! We get

$$P(r) = \rho_0 c^2 \left(\frac{(1-2GMr^2/R^3c^2)^{1/2} - (1-2GM/Rc^2)^{1/2}}{3\left(1-2GM/Rc^2\right)^{1/2} - \left(1-2GMr^2/R^3c^2\right)^{1/2}}\right) \tag{4.267}$$

$$P_c^{\text{relativistic}} = \rho_0 c^2 \frac{1 - \sqrt{1 - \frac{2GM}{Rc^2}}}{3\sqrt{1 - \frac{2GM}{Rc^2}} - 1} \,. \tag{4.268}$$

In the classical model we had a finite central pressure for each value of the mass and density; now, instead, in order for the pressure to not diverge we must require

$$R > \frac{9}{8} \frac{2GM}{c^2} \,. \tag{4.269}$$

This is known as the **Buchdahl bound**; the radius of a non-black-hole (with constant density) cannot be arbitrarily close to the Schwarzschild radius, it must be at least 12.5 % larger in order for the pressure to not diverge at the center.[9]

This also yields a mass limit for neutron stars. Let us estimate the constant density ρ_0 by assuming that each neutron takes up a sphere of radius its Compton wavelength: $r_n \approx h/m_n c$, so

$$\rho_0 \approx \frac{m_n}{\frac{4\pi}{3} r_n^3} \approx \frac{3 m_n^4 c^3}{4\pi h^3} \,. \tag{4.270}$$

A more accurate estimate would be given by $r_n \approx 0.7 h/m_n c$.

so that, using the fact that $\rho_0 = M/(4\pi R^3/3)$, we can start manipulating the Buchdahl bound:

$$M < \frac{4c^2}{9GR} = \frac{4c^2}{9G} \left(\frac{4\pi\rho_0}{3M} \right)^{1/3} \tag{4.271}$$

$$M^{2/3} < \frac{4c^2}{9G} \left(\frac{4\pi}{3} \frac{3 m_n^4 c^3}{4\pi h^3} \right)^{1/3} \tag{4.272}$$

$$M < \left(\frac{8\pi}{9} \right)^{3/2} M_* \,. \tag{4.273}$$

This yields a bound on the order of $M \lesssim 5 M_\odot$.

[9] More stringent bounds can also be derived — in Lattimer and Prakash [LP07, fig. 2] a plot is shown of possible equations of state of neutron stars in a mass versus radius plane. The bound we derived is denoted as $P < \infty$, and we can also see a "causality" bound, which is related to the speed of sound in an ultrarelativistic medium. Realistic equations of state can reach approximately $R \gtrsim 3GM/c^2$.

Chapter 5

Structure formation

5.1 The nonlinear evolution of a spherical perturbation

We want to discuss how a dark matter halo might form and evolve. In order to do so, we must make certain simplifying assumptions about the shape of the perturbation we want to consider. We will assume spherical symmetry: the fact that this is reasonable is not obvious, and was the subject of debate historically; the American school used spherical models, while the Russian school studied "pancakes", ellipsoids for which one axis was much shorter than the others.

A result which can be derived is that, starting from a generic ellipsoid, if it is less dense than the background it will tend towards a sphere, while it will become less spherical if it is denser than the background.

So, our model being over-dense and spherical will be kind of unphysical, right?

We will discuss the evolution of a spherical perturbation in the shape of a **top-hat**: a constant-density spherical region, embedded in a universe whose density is constant as well, and which is described by the usual FLRW metric, assuming zero spatial curvature and matter dominance.

Let us denote the background density as $\rho_b(t)$; by the results we know about the Einstein-De Sitter model this will be given by $\rho_b(t) = 1/(6\pi G t^2)$.

In the perturbed region we will have a different density, $\rho(\vec{x}, t)$. As we did earlier, when discussing gravitational collapse, we introduce the dimensionless density perturbation

$$\delta(\vec{x}, t) = \frac{\rho(\vec{x}, t) - \rho_b(t)}{\rho_b(t)} . \tag{5.1}$$

This can have values from -1 to $+\infty$; when it is positive we have an *over-density* while when it is negative we have an *under-density*. We will assume that $0 < \delta \ll 1$: a small over-density, which will allow us to apply perturbation theory.

Earlier we found that if we take the linear order of the equations of motion for a perturbation in an Einstein-De Sitter universe we get a growing mode $\delta \propto t^{2/3}$ and a decaying mode $\delta \propto t^{-1}$, for which the velocities read $v \propto t^{1/3}$ and $v \propto t^{-4/3}$ respectively.

If t_i is the initial time, then the density perturbation at a time t is given by

$$\delta(t) = \delta_+(t_i)\left(\frac{t}{t_i}\right)^{2/3} + \delta_-(t_i)\left(\frac{t}{t_i}\right)^{-1}. \tag{5.2}$$

The linearized continuity equation (4.110) allows us to express the velocity in terms of the derivative of the density:

$$v = i\frac{\dot{\delta}}{k}a \propto \left(\frac{2}{3}\delta_+(t_i)\left(\frac{t}{t_i}\right)^{1/3} - \delta_-(t_i)\left(\frac{t}{t_i}\right)^{-4/3}\right), \tag{5.3}$$

since $a \propto t^{2/3}$.

We suppose that at $t = t_i$ we have *unperturbed Hubble flow* — the comoving coordinates of each particle are constant, so $v(t_i) = 0$. Imposing this for the equation we just found for the velocity we get

$$\delta_-(t_i) = \frac{2}{3}\delta_+(t_i)\,, \tag{5.4}$$

so we can express the initial density as

$$\delta(t_i) = \delta_i = \delta_+ + \delta_- = \frac{5}{3}\delta_+\,. \tag{5.5}$$

Our perturbation can be dealt with as if it were a *local FRLW closed universe*: if the background universe is flat, with $k = 0$, then $\Omega_{\mathrm{bg}} = 1$, so in the perturbed "bubble" the perturbed density parameter will be

$$\Omega_p(t_i) = 1 + \delta_i > 1\,. \tag{5.6}$$

When studying curved models, we have shown that they exhibit a *turnaround time* after which the scale factor decreases: since our bubble behaves like a closed Einsten-De Sitter universe, it will do the same. After the turnaround, a closed universe collapses to a single point at a cosmic time $t_{\mathrm{collapse}} = 2t_{\mathrm{turnaround}}$. This will not actually happen in our perturbed model: as the cloud nears collapse oscillations dissipate energy, so that the equilibrium configuration is a finitely-dense cloud with a radius R_{vir}, where "vir" stands for "virialized". These effects take over at the very end of the collapse, so we can estimate the time until equilibrium is reached as twice the turnaround time.

A corollary of Birkhoff's theorem tells us that we can then treat this region using the Friedmann equations with $k = +1$: the first FE reads

$$\dot{a}^2 = \frac{8\pi G}{3}\rho a^2 - k\,, \tag{5.7}$$

which we can write as

$$-k = (1 - \Omega_p)a^2 H^2\,. \tag{5.8}$$

This allows us to write the following equation by substituting this equation evaluated at the initial time t_i: $-k = (1-\Omega_p(t_i))a_i^2 H_i^2$:

$$\dot{a}^2 = H^2\Omega_p a^2 + (1-\Omega_p(t_i))a_i^2 H_i^2 \tag{5.9}$$

$$\frac{\dot{a}^2}{a_i^2} = H^2\Omega_p \frac{a^2}{a_i^2} + (1-\Omega_p(t_i))H_i^2 \tag{5.10}$$

$$\frac{\dot{a}^2}{a_i^2} = H_i^2\left(\Omega_p(t_i)\frac{a_i}{a} + (1-\Omega_p(t_i))\right), \tag{5.11}$$

where we introduce the index p to denote the fact that we are talking about a perturbation.

The calculation does not seem to work out... Is the factor a_i/a right? The first two equations are what I'd do, the last is what Pacciani writes.

The perturbed density $\rho_p(t)$ evolves like

$$\rho_p(t) = \rho_p(t_i)\left(\frac{a_p(t_i)}{a_p(t)}\right)^3 \tag{5.12a}$$

$$= \rho_b(t_i)\Omega_p(t_i)\left(\frac{a_p(t_i)}{a_p(t)}\right)^3. \tag{5.12b}$$

It can be shown that, if we choose the turnaround time t_m by imposing $\dot{a}(t_m) = 0$, we find a density equal to

How does the calculation actually go? It is not clear to me how to get to this next equation

$$\rho_p(t_m) = \rho_b(t_i)\Omega_p(t_i)\left(\frac{\Omega_p(t_i)-1}{\Omega_p(t_i)}\right)^3 \tag{5.13a}$$

$$= \rho_b(t_i)\frac{\left(\Omega_p(t_i)-1\right)^3}{\Omega_p(t_i)^2}. \tag{5.13b}$$

As we discussed with curved models, the turnaround time can be calculated by finding a parametric solution to the Friedmann equations: this yields, changing the reference time from "0" (now) to the initial moment t_i,

$$t_m = \frac{\pi}{2H_i}\frac{\Omega_i}{(\Omega_i-1)^{3/2}} = \frac{\pi}{2H_i}\left(\frac{\rho_b(t_i)}{\rho_p(t_m)}\right)^{1/2}, \tag{5.14}$$

where we inserted the expression we found for the density calculated at the turnaround moment.

Since we assumed that at the initial time there was unperturbed Hubble flow, at that time the Hubble parameter inside and outside was the same, and we can compute it through the first Friedmann equation applied to the background spacetime:

$$H^2(t_i) = \frac{8\pi G}{3}\rho_b(t_i), \tag{5.15}$$

so we have a cancellation, since the ratio $\rho_b(t_i)^{1/2}/H_i$ yields a constant:

$$t_m = \frac{\pi}{2H_i}\left(\frac{\rho_b(t_i)}{\rho_p(t_m)}\right)^{1/2} = \left(\frac{3\pi}{32G\rho_p(t_m)}\right)^{1/2} \tag{5.16}$$

$$\rho_p(t_m) = \frac{3\pi}{32Gt_m^2}\,, \tag{5.17}$$

which holds inside the bubble.

Outside it, the density evolves according to the usual law:

$$\rho_b(t_m) = \frac{1}{6\pi G t_m^2}\,. \tag{5.18}$$

We are implicitly using the *synchronous gauge*, in which the proper time defines the time coordinate for each observer. The exact solution to the Einstein Field Equations we found is known as the Lemaître-Tolman-Bondi solution.

We can then ask: at the turnaround time, how much is the interior density larger than the exterior one? this is given by

$$1 + \delta_p(t_m) = \chi(t_m) = \frac{\rho_p(t_m)}{\rho_b(t_m)} = \frac{3\pi}{32G}6\pi G = \left(\frac{3\pi}{4}\right)^2 \approx 5.6\,. \tag{5.19}$$

This means that $\delta_p(t_m) \approx 4.6$. This certainly is not smaller than 1, so it makes sense to have sought an exact solution instead of a perturbative one.

How would linear theory have fared? It would *not* have predicted a turnaround: in it the growing mode keeps growing, so in order to make the comparison we will need use the expression for the turnaround time from the exact model.

The density perturbation at the turnaround time, considering only the growing mode since the decaying one becomes negligible, would have been

$$\delta_p(t_m) \approx \delta_+(t_i)\left(\frac{t_m}{t_i}\right)^{2/3} = \frac{3}{5}\delta_p(t_i)\left(\frac{t_m}{t_i}\right)^{2/3}\,. \tag{5.20}$$

Let us take the expression for the turnaround time and substitute $H_i = 2/(3t_i)$; also, we want to express the turnaround time in terms of the density perturbation, so we must use $\Omega_p - 1 = (1+\delta_p) - 1 = \delta_p$. This yields

$$t_m = \frac{3\pi t_i}{4} \times \left(\frac{1+\delta_i}{\delta_i^{3/2}}\right) \approx \frac{3\pi t_i}{4}\delta_i^{-3/2}\,, \tag{5.21}$$

since δ_i is small by assumption.

The ratio t_m/t_i, which appears in the expression for the perturbation at the turnaround time, is then given by

$$\frac{t_m}{t_i} = \frac{3\pi}{4}\delta_i^{-3/2}\,, \tag{5.22}$$

which fortunately means that the size of the initial perturbation cancels out:

$$\delta_p(t_m) = \frac{3}{5}\delta_i \left(\frac{3\pi}{4} \delta_i^{-3/2} \right)^{2/3} = \frac{3}{5} \left(\frac{3\pi}{4} \right)^{2/3} \approx 1.06 \,. \tag{5.23}$$

As expected, when the perturbation grows linear theory stops giving us a good approximation for the result.

The virialization time As mentioned before, after the *virialization time* the perturbation has become a halo with a higher temperature than before, and for which the virial theorem applies, and we estimate it as

$$t_{\text{vir}} \approx t_{\text{collapse}} = 2t_m \,. \tag{5.24}$$

The virial theorem tells us that $2T + E_{\text{gr}} = 0$, so the total energy of the system is given by

$$E_{\text{tot}} = T + E_{\text{gr}} = \frac{1}{2} E_{\text{gr}} = -T \,. \tag{5.25}$$

The total energy at virialization is given by

$$E_{\text{vir}} = -\frac{1}{2} \underbrace{\frac{3}{5} \frac{GM^2}{R_{\text{eq}}}}_{-E_{\text{gr}}} \,, \tag{5.26}$$

where the factor $3/5$ comes from the spherical symmetry of the system — we are calculating the gravitational potential energy of a constant-density sphere of radius R_{eq}.

At the time of collapse, instead, there is no kinetic energy since we are at a stationary point, so we only have the gravitational contribution:

$$E_m = -\frac{3}{5} \frac{GM^2}{R_m} \,. \tag{5.27}$$

Energy conservation tells us that $E_m = E_{\text{vir}}$, which means that the radius at virialization, R_{eq}, is twice the radius at the start of the collapse, R_m: $2R_{\text{eq}} = R_m$, since the factor $1/2$ and the radius are the only difference between the formulas.

Since the mass is the same while the volume shrinks by a factor $8 = 2^3$, this then means $\rho_{\text{vir}} = 8\rho_m$.

With this, we can calculate the ratio of the density of the virialized cloud to the density of the background: we find

$$\frac{\rho_p(t_c)}{\rho_c(t_c)} = \underbrace{\frac{\rho_p(t_c)}{\rho_p(t_m)}}_{8} \underbrace{\frac{\rho_p(t_m)}{\rho_b(t_m)}}_{\chi \approx 5.6} \underbrace{\frac{\rho_b(t_m)}{\rho_b(t_c)}}_{2^2} \approx 180 \,, \tag{5.28}$$

where the 2^2 factor for the background density comes from the fact that $\rho_b \propto t^{-2}$ and $t_c = 2t_m$.

What would happen if we were to use linear theory at this time? It predicts that the density scales as $t^{2/3}$, so we get

$$\delta_+(t_c) = \delta_+(t_m)\left(\frac{t_c}{t_m}\right)^{2/3} \approx \frac{3}{5}\left(\frac{3\pi}{4}\right)^{2/3} 2^{2/3} \approx 1.686\,. \tag{5.29}$$

Once again, we see that linear theory cannot be applied in this context: it is not able to predict the large over-densities which are generated by the collapse of clouds.

This result might seem trivial: we already knew that linear theory did not apply! However, this is in fact very useful: in many contexts linear theory can still be applied successfully, but it can be hard to know when the approximation it provides breaks down. This value is then used as a heuristic: when linear theory predicts an over-density of $\delta \sim 1.686$ we know that what physically would have happened there is a collapse with $\delta \sim 180$.

5.2 Press-Schechter theory

This is a theory which was developed in the seventies, and which allows us to study structure formation by estimating the *mass function*

$$n(M) = \frac{\mathrm{d}N}{\mathrm{d}M} = \#\text{ of objects per unit volume with mass in } [M, M+\mathrm{d}M]\,, \tag{5.30}$$

where "objects" is taken to mean "virialized clouds", described according to the top-hat collapse of the last section.

We will use linear theory and apply the "$\delta(\vec{x}, t) > 1.686$ criterion". Let us denote this critical perturbation value as $\delta_c = 1.686$.

In order to only consider objects above a certain mass, we use spatial filtering: we ignore perturbations whose characteristic length is higher than a certain radius R, which corresponds to a certain mass $M \propto R^3$. This is obtained by the application of a low-pass filter $W_R(\vec{x})$.

The probability density of seeing a perturbation δ_M (the index M denotes the fact that we applied the low-pass filter) is well approximated by a Gaussian

$$p(\delta_M)\,\mathrm{d}\delta_M = \frac{1}{\sqrt{2\pi\sigma_M^2}} \exp\left(-\frac{\delta_M^2}{2\sigma_M^2}\right) \mathrm{d}\delta_M\;, \tag{5.31}$$

where the variance is typically diverging before the application of the filter; after its application instead

$$\sigma_M^2 = \left\langle \delta_M^2 \right\rangle \propto M^{-2\alpha}\,, \tag{5.32}$$

and typically $\alpha \sim 1/2$.

In order to have nonvanishing probabilities of $\delta > \delta_c$ we must have σ be quite large, almost of order one: how do we deal with the tail $\delta < -1$ then?

We can then calculate the probability of seeing a perturbation higher than δ_c in a generic location as

$$\mathbb{P}_{>\delta_c}(M) = \int_{\delta_c}^{\infty} \mathrm{d}\delta_M \, p(\delta_M) \,. \tag{5.33}$$

This basically gives us the fraction of the universe which is occupied by virialized objects of mass smaller than M.

Also counting their "areas of influence", right? the region which has $\delta > \delta_c$ in linear theory is much larger than the true volume of the perturbation...

$$\begin{aligned} n(M)M\,\mathrm{d}M &= \rho_m \big(\mathbb{P}_{>\delta_c}(M) - \mathbb{P}_{>\delta_c}(M + \mathrm{d}M)\big) && (5.34\text{a}) \\ &= \rho_m \left| \frac{\mathrm{d}\mathbb{P}_{>\delta_c}}{\mathrm{d}M} \right| \mathrm{d}M && (5.34\text{b}) \\ &= \rho_m \left| \frac{\mathrm{d}\mathbb{P}_{>\delta_c}}{\mathrm{d}\sigma_M} \right| \left| \frac{\mathrm{d}\sigma_M}{\mathrm{d}M} \right| \mathrm{d}M \,. && (5.34\text{c}) \end{aligned}$$

Integrating this expression from $M = 0$ to $M \to \infty$ we expect to find ρ_m, but if we actually compute it we find $\rho_m/2$.

This comes from a miscount: as the mass we are considering shrinks, we might be already including smaller objects inside the gravitational influence of larger ones. Properly accounting for this one gets precisely a factor 2; this means that all the matter in the universe is found in virialized objects.

Integrating, then, with the factor 2 and taking $\sigma_M = (M/M_0)^{-\alpha}$ we find

$$n(M) = \sqrt{\frac{2}{\pi}} \frac{\rho_m}{M_*^2} \alpha \left(\frac{M}{M_*} \right)^{\alpha - 2} \exp\left(- \left(\frac{M}{M_*} \right)^{2\alpha} \right) , \tag{5.35}$$

where $M_* = (2/\delta_c)^{1/2\alpha} M_0$.

Usually the application this theory finds is the calculation of the density of dark matter halos, but in principle it could also be used in order to find the luminosity function of galaxies. Surely this function resembles the Schechter luminosity function $\Phi(L)$ if we set $\alpha = 1/2$ and assume $M/L = \text{const}$. This assumption is, however, not very accurate, so empirical estimates of the luminosity function are preferred when available.

In 1999 Sheth, Mo, and Tormen [SMT01] improved the estimates significantly by accounting for nonspherical collapse — specifically, they allowed for halos shaped like ellipsoids with arbitrary axes. Their results very closely resemble those coming from N-body simulations.

Bibliography

[ABG48] R. A. Alpher, H. Bethe, and G. Gamow. "The Origin of Chemical Elements". In: *Physical Review* 73.7 (Apr. 1, 1948), pp. 803–804. DOI: 10.1103/PhysRev.73.803. URL: https://link.aps.org/doi/10.1103/PhysRev.73.803 (visited on 2020-03-17) (cit. on p. 49).

[And10] S. Andreon. "The Stellar Mass Fraction and Baryon Content of Galaxy Clusters and Groups". In: *Monthly Notices of the Royal Astronomical Society* 407.1 (Sept. 1, 2010). Comment: MNRAS, in press, pp. 263–276. ISSN: 00358711. DOI: 10.1111/j.1365-2966.2010.16856.x. arXiv: 1004.2785. URL: http://arxiv.org/abs/1004.2785 (visited on 2020-09-14) (cit. on p. 17).

[Aut18] O'Raifeartaigh Cormac author Aut. "Investigating the Legend of Einstein's "Biggest Blunder"". In: (Oct. 30, 2018). DOI: 10.1063/PT.6.3.20181030a. URL: https://physicstoday.scitation.org/do/10.1063/PT.6.3.20181030a/abs/ (visited on 2020-03-11) (cit. on p. 41).

[BM84] J. Bekenstein and M. Milgrom. "Does the Missing Mass Problem Signal the Breakdown of Newtonian Gravity?" In: *The Astrophysical Journal* 286 (Nov. 1, 1984), pp. 7–14. ISSN: 0004-637X. DOI: 10.1086/162570. URL: http://adsabs.harvard.edu/abs/1984ApJ...286....7B (visited on 2020-03-03) (cit. on p. 14).

[Bre+02] G. Bressi et al. "Measurement of the Casimir Force between Parallel Metallic Surfaces". In: *Physical Review Letters* 88.4 (Jan. 15, 2002). Comment: 4 Figures, p. 041804. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.88.041804. arXiv: quant-ph/0203002. URL: http://arxiv.org/abs/quant-ph/0203002 (visited on 2020-09-15) (cit. on p. 71).

[CDG98] A. G. Cohen, A. De Rujula, and S. L. Glashow. "A Matter-Antimatter Universe?" In: *The Astrophysical Journal* 495.2 (Mar. 10, 1998). Comment: 28 pages, 5 figures, uses graphicx package; updated for Ap. J. version. Includes additional discussion of turbulence, pp. 539–549. ISSN: 0004-637X, 1538-4357. DOI: 10.1086/305328. arXiv: astro-ph/9707087. URL: http://arxiv.org/abs/astro-ph/9707087 (visited on 2020-09-16) (cit. on p. 75).

[Col16] Planck Collaboration. "Planck 2015 results. XIII. Cosmological parameters". In: *Astronomy & Astrophysics* 594 (Sept. 2016), A13. ISSN: 1432-0746. DOI: 10.1051/0004-6361/201525830. URL: http://dx.doi.org/10.1051/0004-6361/201525830 (cit. on pp. 5, 18, 25, 33, 35, 50, 66).

[Col18a] Planck Collaboration. *Planck 2018 results. I. Overview and the cosmological legacy of Planck*. 2018. arXiv: 1807.06205 [astro-ph.CO] (cit. on pp. 4, 15).

[Col18b] The Fermi-LAT Collaboration. “A Gamma-Ray Determination of the Universe’s Star Formation History”. In: *Science* 362.6418 (Nov. 30, 2018), pp. 1031–1034. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aat8123. pmid: 30498122. URL: https://science.sciencemag.org/content/362/6418/1031 (visited on 2020-07-30) (cit. on p. 17).

[DL04] Tamara M. Davis and Charles H. Lineweaver. “Expanding Confusion: Common Misconceptions of Cosmological Horizons and the Superluminal Expansion of the Universe”. In: *Publications of the Astronomical Society of Australia* 21.1 (2004). Comment: To appear in Publications of the Astronomical Society of Australia, 26 pages (preprint format), 6 figures. Version 2: Section 4.1 revised, pp. 97–109. ISSN: 1323-3580, 1448-6083. DOI: 10.1071/AS03040. arXiv: astro-ph/0310808. URL: http://arxiv.org/abs/astro-ph/0310808 (visited on 2020-03-09) (cit. on p. 25).

[dGra+19] Anna de Graaff et al. “Probing the Missing Baryons with the Sunyaev-Zel’dovich Effect from Filaments”. Version 1. In: *Astronomy & Astrophysics* 624 (Apr. 2019). Comment: 13 pages, 8 figures; accepted for publication in A&A, A48. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/201935159. arXiv: 1709.10378. URL: http://arxiv.org/abs/1709.10378 (visited on 2020-07-30) (cit. on p. 17).

[Ein22] A. Einstein. *Comment on A. Friedmann's Paper "On the Curvature of Space"*. Letter. Einstein wrongly corrects Friedmann. 1922. URL: https://einsteinpapers.press.princeton.edu/vol13-trans/301 (visited on 2020-03-11) (cit. on p. 39).

[FGN98] Valerio Faraoni, Edgard Gunzig, and Pasquale Nardone. “Conformal Transformations in Classical Gravitational Theories and in Cosmology”. In: (Nov. 14, 1998). Comment: LaTeX, 54 pages, no figures. To appear in Fundamentals of Cosmic Physics. arXiv: gr-qc/9811047. URL: http://arxiv.org/abs/gr-qc/9811047 (visited on 2020-09-16) (cit. on p. 73).

[Few95] M. P. Fewell. “The Atomic Nuclide with the Highest Mean Binding Energy”. In: *American Journal of Physics* 63.7 (July 1, 1995), pp. 653–658. ISSN: 0002-9505. DOI: 10.1119/1.17828. URL: https://aapt.scitation.org/doi/10.1119/1.17828 (visited on 2020-09-24) (cit. on p. 121).

[Fix09] D. J. Fixsen. “The Temperature of the Cosmic Microwave Background”. In: *The Astrophysical Journal* 707.2 (Nov. 2009), pp. 916–920. ISSN: 1538-4357. DOI: 10.1088/0004-637x/707/2/916. URL: http://dx.doi.org/10.1088/0004-637X/707/2/916 (cit. on p. 4).

[Fri22] A. Friedmann. *From Alexander Friedmann*. Letter. 1922. URL: https://einsteinpapers.press.princeton.edu/vol13-trans/363 (visited on 2020-03-11) (cit. on p. 39).

[Har93] E. Harrison. “The Redshift-Distance and Velocity-Distance Laws”. In: (1993). URL: http://adsabs.harvard.edu/full/1993ApJ...403...28H (visited on 2020-03-24) (cit. on p. 25).

[Hog00] David W. Hogg. "Distance Measures in Cosmology". Version 4. In: (Dec. 15, 2000). Comment: This short and purely pedagogical text is not submitted anywhere but here. Errors reported to the author will be rewarded with gratitude and acknowledgements. arXiv: astro-ph/9905116. URL: http://arxiv.org/abs/astro-ph/9905116 (visited on 2020-03-07) (cit. on pp. 25, 39).

[Hos19] S. Hossenfelder. "Screams for Explanation: Finetuning and Naturalness in the Foundations of Physics". In: *Synthese* (Sept. 3, 2019). Comment: 17 pages, no figures, typos fixed. ISSN: 0039-7857, 1573-0964. DOI: 10.1007/s11229-019-02377-5. arXiv: 1801.02176. URL: http://arxiv.org/abs/1801.02176 (visited on 2020-09-14) (cit. on p. 70).

[Hub29] Edwin Hubble. "A Relation Between Distance And Radial Velocity Among Extra-Galactic Nebulae". In: (1929). URL: https://www.pnas.org/content/pnas/15/3/168.full.pdf (visited on 2020-03-03) (cit. on p. 19).

[Hus16] Lars Husdal. "On Effective Degrees of Freedom in the Early Universe". In: *Galaxies* 4.4 (Dec. 17, 2016). Comment: 29 pages, 7 figures, p. 78. ISSN: 2075-4434. DOI: 10.3390/galaxies4040078. arXiv: 1609.04979. URL: http://arxiv.org/abs/1609.04979 (visited on 2020-06-30) (cit. on p. 61).

[JR05] Nils Voje Johansen and Finn Ravndal. "On the Discovery of Birkhoff's Theorem". Version 2. In: (Sept. 6, 2005). Comment: 4 pages, no figures. More references have been included, small changes in text plus new data on Eiesland. arXiv: physics/0508163. URL: http://arxiv.org/abs/physics/0508163 (visited on 2020-03-08) (cit. on p. 27).

[Kee14] Charles Keeton. "Star and Planet Formation". In: *Principles of Astrophysics: Using Gravity and Stellar Physics to Explore the Cosmos*. Ed. by Charles Keeton. Undergraduate Lecture Notes in Physics. New York, NY: Springer, 2014, pp. 377–394. ISBN: 978-1-4614-9236-8. DOI: 10.1007/978-1-4614-9236-8_19. URL: https://doi.org/10.1007/978-1-4614-9236-8_19 (visited on 2020-04-21) (cit. on p. 101).

[LP07] James M. Lattimer and Maddapa Prakash. "Neutron Star Observations: Prognosis for Equation of State Constraints". In: *Physics Reports* 442.1-6 (Apr. 2007). Comment: 70 pages, 19 figures, submitted to Hans Bethe Centennial Physics Reports, pp. 109–165. ISSN: 03701573. DOI: 10.1016/j.physrep.2007.02.003. arXiv: astro-ph/0612440. URL: http://arxiv.org/abs/astro-ph/0612440 (visited on 2020-09-26) (cit. on p. 140).

[LC02] F. Lucchin and P. Coles. *Cosmology*. 2nd ed. Wiley, 2002 (cit. on pp. 3, 36, 75, 79, 80, 83, 84).

[Lud99] Malcolm Ludvigsen. *General Relativity: A Geometric Approach*. For distinction between static and stationary spacetimes. Cambridge University Press, May 28, 1999. 234 pp. ISBN: 978-0-521-63976-7. Google Books: YA8rxOn9H1sC (cit. on p. 39).

[MTW73] C.W. Misner, K.S. Thorne, and J.A. Wheeler. *Gravitation*. W.H. Freeman & Co., 1973 (cit. on p. 27).

[NS19] Jose Natario and Amol Sasane. "Decay of Solutions to the Klein-Gordon Equation on Some Expanding Cosmological Spacetimes". In: (Sept. 9, 2019). Comment: 52 pages, 4 figures. arXiv: 1909.01292 [gr-qc, physics:math-ph]. URL: http://arxiv.org/abs/1909.01292 (visited on 2020-09-16) (cit. on p. 74).

[Pac18a] L. Pacciani. *Appunti Del Corso The Physical Universe*. Pacciani notes for Fundamentals of Astrophysics and Cosmology. 2018. URL: https://leonardo.pm/teaching/ (visited on 2020-03-03) (cit. on pp. 70, 79).

[Pac18b] L. Pacciani. *Appunti del corso The Physical Universe*. 2018. URL: https://leonardo.pm/teaching/#other-documents (cit. on pp. 3, 84, 86, 91, 93, 105).

[Pla+19] Planck Collaboration et al. "Planck 2018 Results. I. Overview and the Cosmological Legacy of Planck". In: (Dec. 3, 2019). arXiv: 1807.06205 [astro-ph]. URL: http://arxiv.org/abs/1807.06205 (visited on 2020-03-03) (cit. on p. 18).

[Rie+16] Adam G. Riess et al. "A 2.4% Determination of the Local Value of the Hubble Constant". In: *The Astrophysical Journal* 826.1 (July 21, 2016). Comment: accepted ApJ, includes proof corrections and edits, 63 pages, 16 figures, 8 tables. Table 4 available electronically by ApJ Revised since v1 to include one new supernova/calibrator and updated Planck constraints, p. 56. ISSN: 1538-4357. DOI: 10.3847/0004-637X/826/1/56. arXiv: 1604.01424. URL: http://arxiv.org/abs/1604.01424 (visited on 2020-03-05) (cit. on p. 18).

[SMT01] Ravi K. Sheth, H. J. Mo, and Giuseppe Tormen. "Ellipsoidal Collapse and an Improved Model for the Number and Spatial Distribution of Dark Matter Haloes". In: *Monthly Notices of the Royal Astronomical Society* 323.1 (May 1, 2001). Comment: 12 pages, 6 figures, submitted to MNRAS, pp. 1–12. ISSN: 0035-8711, 1365-2966. DOI: 10.1046/j.1365-8711.2001.04006.x. arXiv: astro-ph/9907024. URL: http://arxiv.org/abs/astro-ph/9907024 (visited on 2020-09-27) (cit. on p. 147).

[TM20] J. Tissino and G. Mentasti. *General Relativity Notes*. 2020. URL: https://github.com/jacopok/notes/blob/master/ap_first_semester/general_relativity/main.pdf (cit. on pp. 27, 32).

[Toj] Rita Tojeiro. "Understanding the Cosmic Microwave Background Temperature Power Spectrum". In: (), p. 9 (cit. on p. 66).

[WA03] H. J. Weber and G. B. Afken. *Essential Mathematical Methods for Physicists*. Academic Press, 2003 (cit. on p. 60).

[Wei72] S. Weinberg. *Gravitation and Cosmology - Principles and Applications of the General Theory of Relativity*. Wiley, 1972. ISBN: 0-471-92567-5 (cit. on pp. 50, 54).

[Wol] Natalie Wolchover. *Neutron Lifetime Puzzle Deepens, but No Dark Matter Seen*. URL: https://www.quantamagazine.org/neutron-lifetime-puzzle-deepens-but-no-dark-matter-seen-20180213/ (visited on 2020-09-19) (cit. on p. 86).

[Won+19] Kenneth C. Wong et al. *H0LiCOW XIII. A 2.4% measurement of H_0 from lensed quasars: 5.3σ tension between early and late-Universe probes*. 2019. arXiv: 1907.04869 [astro-ph.CO] (cit. on p. 11).

[Wri03] E. L. Wright. *Theoretical Overview of Cosmic Microwave Background Anisotropy*. 2003. arXiv: astro-ph/0305591 [astro-ph] (cit. on p. 4).

[YBK10] Jaswant K. Yadav, J. S. Bagla, and Nishikanta Khandai. "Fractal dimension as a measure of the scale of homogeneity". In: *Monthly Notices of the Royal Astronomical Society* 405.3 (June 2010), pp. 2009–2015. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2010.16612.x. eprint: https://academic.oup.com/mnras/article-pdf/405/3/2009/2899165/mnras0405-2009.pdf. URL: https://doi.org/10.1111/j.1365-2966.2010.16612.x (cit. on p. 5).

[Yak+] D G Yakovlev et al. "Neutrino Emission from Neutron Stars". In: (), p. 165 (cit. on p. 137).

[ZY12] Shuang-Nan Zhang and Shuxu Yi. "ON A COMMON MISUNDERSTANDING OF THE BIRKHOFF THEOREM AND LIGHT DEFLECTION CALCULATION: GENERALIZED SHAPIRO DELAY AND ITS POSSIBLE LABORATORY TEST". In: *International Journal of Modern Physics: Conference Series* 12 (Jan. 2012), pp. 419–430. ISSN: 2010-1945, 2010-1945. DOI: 10.1142/S2010194512006642. URL: https://www.worldscientific.com/doi/abs/10.1142/S2010194512006642 (visited on 2020-03-08) (cit. on p. 27).