Cosmological perturbation theory and structure formation

Vincent Desjacques

Physics department and Asher Space Science Institute, Technion, Haifa 3200003, Israel

March 2019

Contents

Chapter 1

The FLRW universe

Our Universe is far from homogeneous and isotropic except on the largest scales. Fig.**??** shows two slices that extend over 1.5×10^9 light years in a two-degree field of the sky. Each dot is a galaxy. The distribution of galaxies is clearly not random. Galaxies are arranged on the borders of huge holes or voids and form a delicate *cosmic web* of filaments and walls. Superclusters of galaxies are found at the intersection of this filamentary structure.

1.1 The FRW metric

These observational facts and ideas can be formalized with General Relativity (GR), which will be our fiducial theory of gravity. Basic knowledge of GR should be enough to understand the following discussion.

We will work with the metric signature $(-+++)$, such that a 4-vector v is timelike if $v_\mu v^\mu < 0$ and spacelike if $v_\mu v^\mu > 0$. In a Lorentz frame, the scalar product between two 4-vector v and w thus is $(v, w) = g_{\mu\nu} v^\mu w^\nu = -v^0 w^0 + v^1 w^1 + v^2 w^2 + v^3 w^3$. Note also that the 4-velocity u^μ of a (massive) particle satisfies $-(u^0)^2 + |\mathbf{u}|^2 = -1$. Furthermore, we shall work with the natural units $c = \hbar = k_B = 1$.

1.1.1 Spatial metric

Ignoring for the moment inhomogeneities in the matter distribution, our Universe can be described by a 4-dimensional space-time with homogeneous and isotropic spatial sections. This space-time formally is a pseudo-Riemannian manifold $\mathcal{M}$ with metric g, and is denoted $(\mathcal{M}, g)$. This homogeneous and isotropic space-time admits a slicing into maximally, symmetric 3-spaces More precisely, there is a preferred geodesic time coordinate t, called 'cosmic time', such that the 3-spaces of constant t are maximally symmetric spaces or, equivalently, spaces of constant curvature.

The metric g is therefore of the form

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = -dt^2 + a^2(t) \gamma_{ij} dx^i dx^j \ . \tag{1.1}$$

Here and henceforth, greek indices run over the four space-time dimensions whereas latin indices only run over the spatial dimensions. We will often use spherical coordinates. The corresponding transformation laws are $x^1 = \chi \sin\theta \cos\phi$, $x^2 = \chi \sin\theta \sin\phi$ and $x^3 = \chi \cos\theta$, where χ is the comoving radial coordinate. The geometrical properties of the spatial part $\gamma_{ij} dx^i dx^j$ of the metric depend on the sign of its curvature K. The only simply connected, isotropic and homogeneous 3-dimensional manifolds are the usual 3-dimensional Euclidean space ($K = 0$), the hypersphere ($K > 0$) and the hyperbolic space ($K < 0$). The structure of the three spatial dimensions is frequently used to classify FRW universes, i.e. the universe is *spatially flat* if a slice of constant t is isometric to the Euclidean space, *closed* if isometric to the hypersphere and *open* if isometric to the hyperbolic space.

To unify these three types of universes (flat, closed and open) with a single metric, it is convenient to introduce the generalized trigonometric functions. These are the generalized sine,

$$\sin_K\chi = \begin{cases} K^{-1/2}\sin\left[K^{1/2}\chi\right] & K > 0 \\ \chi & K = 0 \\ (-K)^{-1/2}\sinh\left[(-K)^{1/2}\chi\right] & K < 0 \end{cases} , \tag{1.2}$$

and cosine,

$$\cos_K\chi = \begin{cases} \cos\left[K^{1/2}\chi\right] & K > 0 \\ \chi & K = 0 \\ \cosh\left[(-K)^{1/2}\chi\right] & K < 0 \end{cases} , \tag{1.3}$$

The remaining functions can be determined from the usual trigonometric relations. The comoving radial coordinate χ spans the range $0 \leqslant \chi < \infty$ for spatially flat and open universes, but is restricted to the range $0 \leqslant \chi \leqslant \pi K^{-1/2}$ for spatially closed universes. Note that the exponent of $-1/2$ is introduced for later convenience. With these definitions, all our FRW universes can be described by the metric

$$ds^2 = -dt^2 + a^2(t)\left[d\chi^2 + \sin_K^2\chi\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right] . \tag{1.4}$$

The substitution $r = \sin_K\chi$ yields

$$ds^2 = -dt^2 + a^2(t)\left[\frac{dr^2}{1 - Kr^2} + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right] . \tag{1.5}$$

This form is, however, less convenient for studies of inhomogeneities (CMB anisotropies etc.) due to the complicated radial g_{rr} metric component. Nevertheless, it makes clear that the effect of curvature becomes important only at a comoving radius $r \sim |K|^{-1/2}$. Therefore, we shall define the physical radius of curvature of the universe as $R_{\text{curv}} = a(t)|K|^{-1/2}$.

The metric Eq.(1.4) is also called FLRW metric after Friedmann, Lemaitre, Robertson and Walker.

We can always change the normalization of the scale factor by a suitable choice of the matrix γ_{ij}. One usually assumes that the present value of a is normalized to unity, i.e. $a_0 = a(t_0) = 1$, so that the curvature K is not dimensionless (It has dimension of [length^{-2}]). It is also possible to normalise the scale factor such that $K = \pm 1$ whenever it is different from zero (in which case we denote it k). In this case however, one should bear in mind that a would have the dimension of [length].

1.1.2 Hubble rate and conformal time

As stated above, the scale factor $a(t)$ measures the expansion of the universe, in the sense that the physical separation between two comoving observers is

$$\Delta s(t) = a(t)\sqrt{\gamma_{ij}\Delta x^i \Delta x^j} \propto a(t) . \tag{1.6}$$

Hence, the comoving observers are moving away from (towards) each other when $a(t)$ is increasing (decreasing). The rate at which they approach or recede is given by

$$\frac{d}{dt}\Delta s(t) = \frac{da(t)/dt}{a(t)}\Delta s(t) = H(t)\Delta s(t) \ , \tag{1.7}$$

in agreement with our previous findings. Note that $H(t)$ is insensitive to the normalisation of a.

It is useful to define a quantity η called *conformal time*, given by

$$\eta = \int \frac{dt}{a} = \int \frac{da}{a^2 H(a)} \ . \tag{1.8}$$

We usually set $\eta = 0$ (and for that matter $t = 0$) at the "Big-Bang" $a = 0$. The present-day value of the conformal time is denoted η_0. Substituting η in the FRW metric Eq.(1.4), we obtain

$$ds^2 = a^2(\eta)\left[-d\eta^2 + d\chi^2 + \sin^2_K\chi\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right] \ . \tag{1.9}$$

With this time coordinate, the flat FRW metric is conformal to Minkowski. We can also define a *conformal Hubble rate* $\mathcal{H}$,

$$\mathcal{H} \equiv \frac{da/d\eta}{a} = \frac{da}{dt} = aH \ . \tag{1.10}$$

The conformal time turns out to be particularly useful for the computation of CMB anisotropies. Another advantage is its straightforward relation to causality. Namely, imagine a particle travelling through space-time. We have:

$$\frac{dt_{\text{proper}}}{d\eta} = \sqrt{a^2\left(1 - \gamma_{ij}\frac{dx^i}{d\eta}\frac{dx^j}{d\eta}\right)} = a\sqrt{1 - \left(\frac{ds_C}{d\eta}\right)^2} \equiv \text{real} \tag{1.11}$$

Here, s_C is the comoving distance traveled by the particle. Since the proper time experienced by the particle must be real, then the argument of the square-root must be non-negative, i.e.

$$\frac{ds_C}{d\eta} = \sqrt{\gamma_{ij}\frac{dx^i}{d\eta}\frac{dx^j}{d\eta}} \leqslant 1 \ . \tag{1.12}$$

Therefore, the total comoving distance the particle has traveled since the Big-Bang must be $s_C \leqslant \eta$. In other words, an observer at comoving time η cannot observe any object or process that lies at a comoving distance exceeding η. For this reason, the sphere of comoving radius η surrounding an observer is called the *horizon*.

1.1.3 Redshift

Since we now have a metric for the universe, we can ask about the propagation of particles in it. Consider a photon propagating radially as seen by an observer $\mathcal{O}$, so that the angular coordinates θ and ϕ are fixed in Eq.(1.4). Therefore, the relevant part of the metric is

$$ds^2 = -dt^2 + a^2(t)d\chi^2 \equiv 0 \qquad (\theta, \phi \text{ fixed}) \tag{1.13}$$

The last equality reflects the fact that world lines of light rays are null geodesics. This is equivalent to saying that the propagation of this photon is described by the Lagrangian

$$\mathcal{L} \equiv \frac{1}{2}g_{\mu\nu}\dot{x}^\mu\dot{x}^\mu = \frac{1}{2}\left[-\dot{t}^2 + a^2(t)\dot{\chi}^2\right] = 0 \ , \tag{1.14}$$

where a dot designates a derivative w.r.t. some affine parameter λ. Since $\mathcal{L}$ does not explicitly depend on χ (i.e. χ is a cyclic coordinate), this implies that $\partial\mathcal{L}/\partial\chi = 0$. Therefore, the Euler-Lagrange equation for χ simplifies to

$$\frac{d}{d\lambda}\left(\frac{\partial\mathcal{L}}{\partial\dot{\chi}}\right) - \frac{\partial\mathcal{L}}{\partial\chi} = \frac{d}{d\lambda}\left(\frac{\partial\mathcal{L}}{\partial\dot{\chi}}\right) = 0 \,, \tag{1.15}$$

from which we infer that the (canonical) covariant momentum

$$p_\mu \equiv \frac{\partial\mathcal{L}}{\partial x^\mu} \qquad \Longrightarrow \qquad p_\chi \equiv \frac{\partial\mathcal{L}}{\partial\dot{\chi}} = a^2\dot{\chi} \tag{1.16}$$

is conserved along the photon trajectory. The fact that the radial part of the metric does not depend on χ is a direct consequence of the homogeneity of the universe. Note, however, that p_χ is *not* the physical momentum as measured by a comoving observer. Such an observer will measure an energy

$$E_{\rm obs} = -p_\mu u^\mu \,, \tag{1.17}$$

where $u^\mu = (1,0,0,0)$ is the observer 4-velocity and $p_\mu = (p_0, p_\chi, 0, 0)$ is the photon covariant momentum, whose time component is given by

$$p_0 \equiv \frac{\partial\mathcal{L}}{\partial\dot{t}} = -\dot{t} \,. \tag{1.18}$$

Therefore, the observed energy is

$$E_{\rm obs} = -p_0 u^0 = \dot{t} = a(t)\dot{\chi} = \frac{p_\chi}{a(t)} \,. \tag{1.19}$$

Furthermore, since in the observer frame the relation $E_{\rm obs}^2 - p_{\rm obs}^2 = 0$ (from special relativity) holds, the observed, physical photon momentum is

$$p_{\rm obs} = E_{\rm obs} = a(t)\dot{\chi} = \frac{p_\chi}{a(t)} \,. \tag{1.20}$$

Consequently, both the measured energy $E_{\rm obs}$ and momentum $p_{\rm obs}$ vary inversely with the scale factor a. Moreover, since $E = p_{\rm obs} \propto \lambda^{-1}$, the observer will measure a wavelength $\lambda \propto a$. Therefore, it is convenient to define a time-dependent quantity call the *redshift* by $z = a^{-1} - 1$. The physical momentum of a photon observed today $p_{\rm obs,0}$ is related to its physical momentum $p_{\rm obs,i}$ at some earlier epoch by

$$p_{\rm obs,0} = p_{\rm obs,i}\,\frac{1+z_0}{1+z_i} = \frac{p_{\rm obs,i}}{1+z_i} \tag{1.21}$$

since $z_0 = z(t_0) = 0$ in our convention ($a_0 = 1$). Furthermore, the observed wavelength λ_0 is related to the wavelength λ_i at the time t_i through

$$\lambda_0 = \lambda_i\,(1+z_i) \,. \tag{1.22}$$

As the universe expands, the wavelength of a freely propagating photon increases, just as all physical distances increase with the expansion. The redshift of the photon wavelength is due to the fact that the universe was smaller when it was emitted. Moreover, the scaling $p_{\rm obs} \propto a^{-1}$ also applies for freely propagating massive particles.

Eq.(1.22) is very useful in cosmology since this is essentially the method by which redshifts are actually measured. For instance, if we look at a distant galaxy and observe, e.g., the $1216\AA$ spectral line of hydrogen at a wavelength of 5000 $\AA$, then we immediately conclude that the redshift of the galaxy must be $z = 5000/1216 - 1 = 3.11$.

A few remarks:

- This cosmological redshift has nothing to do with the stars own gravitational field. This contribution is completely negligible compared to the effect of cosmological redshift
- Unlike the gravitational redshift in GR, this cosmological redshift is symmetric between the receiver and the emitter. Namely, light sent from earth to that distant galaxy would likewise be redshifted.
- The cosmological redshift depends only on the value of $a(t)$ at t_i and t_0. Hence, it is fundamentally *not* a cummulative effect caused by the expansion of space, but rather a consequence of the fact that emitter and observer do not share the same intertial frame.
- Astronomers like to define a *recessional velocity* $v = cz$, which is meaningful only if $z \leqslant 1$. However, it is important to bear in mind that the cosmological redshift is fundamentally not a Doppler shift.
- The observed redshift of a galaxy is the sum of its cosmological redshift and a Doppler shift caused by its peculiar velocity relative to the local comoving frame.
- To date, the largest observed redshift of a galaxy is $z \sim 10$, corresponding to a distance of ~ 4000Mpc. The redshift of the CMB is $z \simeq 1000$.

1.1.4 Distances

Since the Universe is expanding, there is no unique way to define distances on our backward light cone. Furthermore, since it is practically impossible to measure physical distance directly, one usually relies on measures of comoving distances such as

- The *line-of-sight comoving distance* D_C.
- The *comoving angular diameter distance* D_{CA}.

These two distances will be discussed in the exercise session.

1.2 Solving the Einstein equations for a FRW universe

1.2.1 Energy-momentum tensor

The energy-momentum tensor can generally be written as

$$T_{\mu\nu} = \rho u_\mu u_\nu + P\big(g_{\mu\nu} + u_\mu u_\nu\big) + \pi_{\mu\nu} \ , \tag{1.23}$$

where

- The *stress tensor* $P\big(g_{\mu\nu} + u_\mu u_\nu\big) + \pi_{\mu\nu}$ encodes the effect of (conservative and non-conservative) forces on the flow.
- The *shear stress* $\pi_{\mu\nu}$ (or anisotropic stress) is symmetric ($\pi_{\mu\nu} = \pi_{\nu\mu}$), transverse ($u^\mu\pi_{\mu\nu} = 0$) and traceless ($\pi^\mu_\mu = 0$). It includes *shear viscosity*.
- u^μ is the fluid 4-velocity, whose components depends on the obsever.
- By contrast, ρ and P are defined in the fluid *rest frame*: ρ is the *energy density* (or often just the "density") and includes both the rest mass and internal energy; P is the *pressure* (which is the sum of the thermodynamic and viscous pressure), or *isotropic stress*.

For a perfect fluid, the stress tensor depends only on the equilibrium thermodynamic pressure. In this case, it does not matter whether the fluid 4-velocity describes energy or particle transport as everything moves the same way. However, this distinction matters for an imperfect fluid. Here and henceforth, we adopt the so-called "Landau frame" in which u^μ describes the transport of energy. Therefore, $j^\mu = \rho u^\mu$ is the energy current and, thereby, includes also heat conduction.

The homogeneity and isotropy of the universe provides two constraints on $T_{\mu\nu}$. First, there is no prefered direction so that $u_\mu = (-1, 0, 0, 0)$ is the fluid 4-velocity. A comoving observer is thus at rest with the fluid. Second, the anisotropic stress must vanish. Therefore, the most general stress-energy tensor consistent with the symmetries of the FRW universe has components

$$T_{00} = \rho \;, \qquad T_{0i} = 0 \;, \qquad T_{ij} = Pa^2\gamma_{ij} \;. \tag{1.24}$$

In general,

- T_{00} is the total energy density
- T_{0i} is the i-component of the energy flux (i.e.. the energy flux through a surface oriented in the i-direction)
- T_{i0} is the 3-momentum density (i.e. the density of momentum component i)
- T_{ij} is the flux of fluid momentum p_i through a surface oriented in the j-direction.

Note that a stress-energy tensor is necessarily symmetric: $T_{0i} = T_{i0}$. Here, the symmetries have left us with two degrees of freedom, ρ and P, which are functions of time solely. P generally includes a viscous pressure, unless the fluid is perfect.

The (rest frame) energy density can generally be defined as the timelike eigenvalue of the stress-energy tensor. Namely, let us consider a *locally inertial frame* such that $g_{\mu\nu}$ reduces to the special relativistic metric $\eta_{\mu\nu}$ at some spacetime point, and a timelike vector u^μ representing the fluid 4-velocity. At that point, we define

$$T^\mu_\nu u^\nu = -\rho u^\mu \;. \tag{1.25}$$

This is a very general definition of the rest frame energy density ρ, which imposes a nontrivial condition on the form of $T_{\mu\nu}$. Indeed, on multiplying the above equality by u_μ, we find

$$T^\mu_\nu u^\nu u_\mu = -\rho u^\mu u_\mu = +\rho \geqslant 0 \tag{1.26}$$

since the energy density of physical (gravitating) matter is always non-negative, and timelike vectors satisfy $u_\mu u^\mu = -1$. The requirement

$$T^\mu_\nu u^\nu u_\mu \equiv T(u,u) \geqslant 0 \,. \tag{1.27}$$

is known as the *weak energy condition*. It implies that the energy density measured by an observer comoving with the fluid is non-negative. The requirement that

$$\left(T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T\right) u^\mu u^\nu \geqslant 0 \tag{1.28}$$

is known as the *strong energy condition*. This inequality implies that gravity is always an attractive force. One would naively expect that only tensors satisfying at least Eq.(1.27) are admissible stress-energy tensors. However, subtle quantum effects or weird classical settings may even lead to the violation of Eq.(1.27). These inequalities predict the occurence of spacetime singularities under very general and physically reasonable conditions.

It is convenient to express the total stress-energy tensor as a sum of the stress-energy tensors of the various components of the universe. For instance, if the universe contains matter (m), radiation (r) and a cosmological constant (Λ), then

$$T_{\mu\nu} = \left[T_m\right]_{\mu\nu} + \left[T_r\right]_{\mu\nu} + \left[T_\Lambda\right]_{\mu\nu} \,. \tag{1.29}$$

Each additional component will add another term to the sum. The total stress-energy tensor is conserved and satisfies the conservation equation $\nabla_\mu T^{\mu\nu} = 0$ (where ∇_ν denotes the covariant derivative w.r.t. the FRW metric Eq.(1.4). In general, the energy-momenta of the various components are not separately conserved because energy and momentum will generally be exchanged among the various components through various interactions (gravity excluded).

1.2.2 Friedmann equations

To obtain a dynamical description of the FRW universe, we can solve Einstein's equations given the FRW metric and the stress-energy tensor Eq.(1.24). We follow the sign convention of Misner, Thorne & Wheeler for $g_{\mu\nu} = (-,+,+,+)$, $R^\sigma_{\mu\nu\lambda} = +...$ and $G_{\mu\nu} = +8\pi G T_{\mu\nu}$. To derive Friedmann's equations, one usually proceeds as follows:

- Read off the $g_{\mu\nu}$ from the FRW metric
- Compute the Christofell symbols $\Gamma^\lambda_{\mu\nu}$ with aid of the relation

$$\Gamma^\lambda_{\mu\nu} = \frac{1}{2} g^{\lambda\sigma} \left(\partial_\nu g_{\mu\sigma} + \partial_\mu g_{\sigma\nu} - \partial_\sigma g_{\mu\nu}\right)$$

  and their symmetry properties $\Gamma^\lambda_{\mu\nu} = \Gamma^\lambda_{\nu\mu}$.

- Successively compute the components of the curvature tensor $R^\sigma_{\mu\nu\lambda}$, the Ricci tensor $R_{\mu\nu}$ as well as the scalar curvature R,

$$R^\sigma_{\lambda\mu\nu} = \partial_\mu \Gamma^\sigma_{\lambda\nu} - \partial_\nu \Gamma^\sigma_{\lambda\mu} + \Gamma^\sigma_{\eta\mu}\Gamma^\eta_{\lambda\nu} - \Gamma^\sigma_{\eta\nu}\Gamma^\eta_{\lambda\mu}$$

$$R_{\mu\nu} = R^\lambda_{\mu\lambda\nu} \qquad \text{and} \qquad R = R^\mu_\mu$$

- Compute the components of the Einstein tensor $G_{\mu\nu}$ and write down Einstein's equations

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu} \ ,$$

  where G is the Newton constant. Note that we have taken into account a cosmological constant Λ.

This task is a bit tedious and not very illuminating (the details can be found in nearly all cosmology textbooks). For sake of illustration, we will consider a second approach based on differential geometry to derive the components of the Einstein tensor $G_{\mu\nu}$.

We begin with the FRW metric in the form

$$ds^2 = -dt^2 + \frac{a^2(t)}{\left(1 + \frac{K}{4} r^2\right)^2} \left[dr^2 + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right] \ , \tag{1.30}$$

where K is proportional to the curvature of the 3-dimensional spatial sections. This form of the FRW metric can be obtained by the replacement $r \to r(1 + \frac{K}{4} r^2)^{-1}$ in the metrix Eq.(1.5). We introduce the orthonormal tetrad θ^μ (i.e. an orthonormal basis of 1-forms),

$$\theta^0 = dt \ , \qquad \theta^i = \frac{a dx^i}{1 + \frac{K}{4} r^2} \equiv \Psi dx^i \ , \tag{1.31}$$

such that $g = \eta_{\mu\nu} \theta^\mu \otimes \theta^\nu$ with $\eta_{\mu\nu} = \mathrm{diag}(-1, 1, 1, 1)$. The 2-forms $d\theta^\mu$ are readily computed as

$$d\theta^0 = 0 \tag{1.32}$$

$$\begin{aligned} d\theta^i &= \frac{\dot{a}}{1 + \frac{K}{4} r^2} dt \wedge dx^i - \frac{a}{\left(1 + \frac{K}{4} r^2\right)^2} \frac{K x_j}{2} dx^j \wedge dx^i \\ &= \frac{\dot{a}}{a} \theta^i \wedge \theta_0 + \frac{K}{2a} x^j \theta^i \wedge \theta_j \ , \end{aligned} \tag{1.33}$$

since $d(r^2) = 2x_j dx^j$, $\theta_0 = \eta_{0\mu}\theta^\mu = -\theta^0$, $\theta_j = +\theta^j$ etc.

We then use the first Cartan's structure equation (which states that there is no torsion)

$$d\theta^\mu + \omega^{\mu\nu} \wedge \theta_\nu = 0 \qquad \Longleftrightarrow \qquad d\theta^\mu = -\omega^{\mu\nu} \wedge \theta_\nu \ , \tag{1.34}$$

where ${\omega^\mu}_\nu$ are the connection 1-forms, to read off the components of $\omega^{\mu\nu} = -\omega^{\nu\mu}$. Clearly, $\omega^{00} = 0$ (anti-symmetry), while the 2-form

$$d\theta^i = \frac{\dot{a}}{a} \theta^i \wedge \theta_0 + \frac{K}{2a} x^j \theta^i \wedge \theta_j \tag{1.35}$$

tells us that

$$\omega^{i0} = -\frac{\dot{a}}{a} \theta^i \qquad \text{and} \qquad \omega^{ij} = \frac{K}{2a}(x^i \theta^j - x^j \theta^i) \ . \tag{1.36}$$

Note that, since $\omega_{\mu\nu}$ is antisymmetric, we have $dg_{\mu\nu} = d\eta_{\mu\nu} = \omega_{\mu\nu} + \omega_{\nu\mu} = 0$. We can now evaluate the 2-forms $d\omega^{\mu\nu}$,

$$d\omega^{i0} = -\frac{\dot{a}}{a} d\theta^i - \partial_k \left(\frac{\dot{a}}{a}\right) dx^k \wedge \theta^i \tag{1.37}$$

$$= -\frac{\ddot{a}}{a} \theta^i \wedge \theta_0 - \frac{K\dot{a}}{2a^2} x^j \theta^i \wedge \theta_j$$

$$d\omega^{ij} = \frac{K}{2a}\left(x^i d\theta^j - x^j d\theta^i\right) + \frac{K}{2}\partial_k \left(\frac{x^i}{a}\right) dx^k \wedge \theta^j - \frac{K}{2}\partial_k \left(\frac{x^j}{a}\right) dx^k \wedge \theta^i \tag{1.38}$$

$$= \frac{K^2}{4a^2} x^l \left(x^i \theta^j - x^j \theta^i\right) \wedge \theta_l + \frac{K}{a^2}\left(1 + \frac{K}{4} r^2\right) \theta^i \wedge \theta^j \ . \tag{1.39}$$

Combining our expressions for $\omega^{\mu\nu}$ and $d\omega^{\mu\nu}$ with the second Cartan's structure equation,

$$d\omega^\mu{}_\nu + \omega^\mu{}_\lambda \wedge \omega^\lambda{}_\nu = \Omega^\mu{}_\nu \tag{1.40}$$

we can easily compute the curvature 2-form Ω^μ_ν and 1-form Ω_μ. Beware of the position of the subscripts and superscripts. We have immediately $\Omega^0{}_0 = 0$. After some algebra, we find

$$\Omega^i{}_0 = \frac{\ddot{a}}{a} \theta^i \wedge \theta_0 \tag{1.41}$$

$$\Omega^i{}_j = \left(\frac{K + \dot{a}^2}{a^2}\right) \theta^i \wedge \theta_j \ . \tag{1.42}$$

The components of the Ricci tensor $R_{\mu\nu}$ can be computed through the relation

$$\Omega_\mu = \mathbf{i}_{\mathbf{e}_\nu} \Omega^\nu{}_\mu = R_{\mu\nu} \theta^\nu \ , \tag{1.43}$$

where $\mathbf{i}_v w$ denotes the interior product of v and w, and $\mathbf{e}_\mu$ is the basis dual to θ^μ. Explicitly,

$$\Omega_0 = \mathbf{i}_{\mathbf{e}_j} \Omega^j{}_0 = 3\frac{\ddot{a}}{a} \theta_0 \tag{1.44}$$

$$\Omega_i = \mathbf{i}_{\mathbf{e}_0} \Omega^0{}_i + \mathbf{i}_{\mathbf{e}_j} \Omega^j{}_i = \left(\frac{\ddot{a}}{a} + 2\frac{K + \dot{a}^2}{a^2}\right) \theta_i \tag{1.45}$$

Therefore, the components of the Ricci tensor in the coordinate basis dx^μ are

$$R_{00} = -3\frac{\ddot{a}}{a} \ , \qquad R_{i0} = 0 \ , \qquad R_{ij} = \left(\frac{\ddot{a}}{a} + 2\frac{K + \dot{a}^2}{a^2}\right) a^2 \gamma_{ij} \tag{1.46}$$

The calculation of the Einstein tensor is now trivial. We leave it as an exercise.

Using Einstein's equations with the stress-energy tensor required to generate a FRW spacetime, we eventually arrive at

$$H^2 + \frac{K}{a^2} = \frac{8\pi G}{3}\rho + \frac{\Lambda}{3} \tag{1.47}$$

$$2\frac{\ddot{a}}{a} + H^2 + \frac{K}{a^2} = -8\pi G P + \Lambda \ . \tag{1.48}$$

It is quite convenient to write the second equation as

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + 3P\right) + \frac{\Lambda}{3} \ . \tag{1.49}$$

These equations are known as the *Friedmann equations*. Note that we can include Λ into ρ and P upon writing $\rho_\Lambda = \Lambda/(8\pi G)$ and $P_\Lambda = -\rho_\Lambda$.

The continuity equation $\nabla_\mu T^{\mu\nu}$ can be useful to find solutions to the Friedmann equations. The spatial components of this relation, $\nabla_\mu T^{\mu i}$, turn out to be identically satisfied since the connection is metric ($\nabla_\mu g_{\nu\lambda} = 0$) and the functions ρ and P depend only on time. We can check this explicitly:

$$\nabla_\mu T^{\mu i} = \nabla_0 T^{0i} + \nabla_j T^{ji} = 0 + P\nabla_j g^{ij} = 0 \ . \tag{1.50}$$

The only interesting conservation law thus is the zero-component $\nabla_\mu T^{\mu 0} = 0$,

$$\nabla_\mu T^{\mu 0} = \partial_\mu T^{\mu 0} + \Gamma^\mu_{\mu\nu} T^{\nu 0} + \Gamma^0_{\mu\nu} T^{\mu\nu} = 0 \tag{1.51}$$

which, for a perfect fluid at rest in the frame (i.e. comoving), becomes

$$\dot\rho + \Gamma^\mu_{\mu 0}\rho + \Gamma^0_{00}\rho + \Gamma^0_{ij} T^{ij} = 0 \ . \tag{1.52}$$

The Christofell symbols $\Gamma^\lambda_{\mu\nu}$ can be easily evaluated from the connection 1-forms $\omega^\mu{}_\nu$ using the relation $\omega^\mu{}_\nu = \Gamma^\mu_{\lambda\nu}\theta^\lambda$. Namely, from $\omega^0{}_0 = 0$ we infer $\Gamma^0_{\lambda 0} = 0$ and, in particular, $\Gamma^0_{00} = 0$. Next, writing $\omega^i{}_0 = \Gamma^i_{\lambda 0}\theta^\lambda$ and comparing with Eq.(1.36) implies that $\Gamma^i_{i0} = 3\dot a/a$. Finally, we have also $\omega^0{}_i = \Gamma^0_{0i}\theta^0 + \Gamma^0_{ji}\theta^j = \Gamma^0_{ji}\theta^j \equiv (\dot a/a)\theta_i$, which leads to $\Gamma^0_{ji} = (\dot a/a) g_{ij}$. Therefore,

$$\Gamma^0_{ij} T^{ij} = \left(\frac{\dot a}{a}\right) P\, g_{ij} g^{ij} = 3HP \ . \tag{1.53}$$

Hence, the continuity equation eventually reads

$$\frac{d\rho}{dt} = -3H\left(\rho + P\right) \ . \tag{1.54}$$

Overall, owing to the Bianchi identities, the Einstein equations and the conservation equations are not independent. In fact, it is easy to see that Eq.(1.47) together with Eq.(1.54) imply the second order equation Eq.(1.49).

To conclude this Section, the Friedmann equations are

$$\mathcal{H}^2 = \frac{8\pi G}{3}\rho a^2 - K \ , \qquad \frac{d\mathcal{H}}{d\eta} = -\frac{4\pi G}{3}\left(\rho + 3P\right) a^2 \ , \tag{1.55}$$

and the continuity equation is

$$\frac{d\rho}{d\eta} = -3\mathcal{H}\left(\rho + P\right) \ . \tag{1.56}$$

when expressed in term of the conformal time η.

1.3 Cosmological models

1.3.1 Critical density and density parameters

Since K is positive, zero or negative depending on the geometry, the first Friedmann equation tells us that the universe is flat if $\rho = \rho_{cr}$, open if $\rho < \rho_{cr}$ and closed if $\rho > \rho_{cr}$, where the *critical density* is

$$\rho_{cr} = \frac{3H^2}{8\pi G} \ . \tag{1.57}$$

The ratio of the total energy density of the universe (including the cosmological constant) to the critical density is denoted as $\Omega = \rho/\rho_{cr}$. Therefore, the universe is open if $\Omega < 1$, flat if $\Omega = 1$ and closed if $\Omega > 1$. We frequently define a density parameter Ω_X for each constituent $X = m, r, \Lambda$ etc. in the universe,

$$\Omega_X = \frac{\rho_X}{\rho_{cr}} \; . \tag{1.58}$$

It is also common to define a curvature parameter $\Omega_K \equiv -K/(aH)^2$ so as to rewrite the first Friedmann equation as

$$\sum_X \Omega_X + \Omega_K = 1 \; . \tag{1.59}$$

Note that Ω_K is not a genuine density, but merely a number to make the Friedmann equation look simple. Beware also that, in the literature, Ω_m, Ω_Λ etc. often refer to their present-day value even though Ω_X generally is redshift-dependent. Here and henceforth, Ω_X will designate the present-day density parameters (so that $\Omega_K = -K/H_0^2$), while we will explicitly write $\Omega_X(z)$ when we refer to their redshift-dependent value. The present-value of the critical density, $\rho_{cr,0} \equiv 3H_0^2/8\pi G$, is

$$\begin{aligned} \rho_{cr,0} &= 1.879 \times 10^{-29}\, h^2 \,\mathrm{g\,cm^{-3}} \\ &= 27.75 \times 10^{10}\ \mathrm{h^2 M_\odot Mpc^{-3}} \\ &= 8.1 \times 10^{-47}\, h^2\ \mathrm{GeV}^4 \; . \end{aligned} \tag{1.60}$$

The second units are frequently used in large scale structure calculations, while the third line is useful for comparison with e.g. phase transitions at high energies.

The present-day values of the density parameters are $\Omega_\Lambda \simeq 0.7$ (i.e. Λ is positive), $\Omega_m \simeq 0.3$ and $\Omega_r \sim 10^{-5}$. Most of the matter today (80%) is in the form of a presumably unkown non-relativistic, *Cold Dark Matter* (CDM) particle whereas the *baryons* which form the stars etc. account only for 20%. Therefore, the matter density parameter is usually split into $\Omega_m = \Omega_c + \Omega_b$, where $\Omega_c \sim 0.25$ and $\Omega_b \sim 0.05$.

The continuity equation is easily solved if the pressure P is given by a linear equation of state of the form $P = w\rho$. In particular, for a (non-relativistic) perfect gas, we have

$$P = \gamma \bar{n} k_B T = \left(\bar{n} m c^2\right)\left(\frac{\gamma k_B T}{mc^2}\right) = (\approx \rho)\left(\frac{c_s}{c}\right)^2 \ll \rho \; , \tag{1.61}$$

where γ is the adiabatic index and c_s the adiabatic speed of sound. Therefore, $P \approx 0$ for non-relativistic matter. On integrating the continuity equation Eq.(1.54) and assuming $w \neq 1$, we arrive at

$$\frac{d\rho}{\rho} = -3(1+w)\frac{da}{a} \quad \Leftrightarrow \quad \rho(a) \propto a^{-3(1+w)} \; . \tag{1.62}$$

Hence, the energy density of relativistic radiation ($w = 1/3$), non-relativistic matter or dust ($w = 0$), and a cosmological constant ($w = -1$) is

$$\rho_r \propto a^{-4} \qquad \text{(radiation)} \tag{1.63}$$

$$\rho_m \propto a^{-3} \qquad \text{(matter)} \tag{1.64}$$

$$\rho_\Lambda = \text{const} \qquad \text{(CC)} \tag{1.65}$$

Note also that, for a thermal blackbody, $\rho_r \propto T^4$ so that $T \propto 1/a$. We will come back to this shortly when we explore the thermal history of the Universe.

Although the cosmological constant dominates today, this was not the case at earlier epochs. In particular, owing to the different scaling with a, we can distinguish three successive *cosmological eras*:

- A *radiation dominated era* (RD) at redshift $z \geqslant z_{\rm eq}$.
- A *matter dominated era* (MD) extending over the range $z_\Lambda \leqslant z \leqslant z_{\rm eq}$
- A *dark-energy dominated era* at redshift $z \leqslant z_\Lambda$

We will precisely work out the *redshift of equivalence* $z_{\rm eq}$ only after discussing photons and neutrinos. In any case, $z_{\rm eq}$ follows from equating the radiation and matter density, $\rho_r = \rho_m$. With the density parameter values quoted above, this is approximately

$$z_{\rm eq} \sim \frac{\Omega_m}{\Omega_r} - 1 \sim 10^4 \,. \tag{1.66}$$

Similarly, z_Λ is obtained from $\rho_m = \rho_\Lambda$, which yields

$$z_\Lambda = (\Omega_\Lambda/\Omega_m)^{1/3} - 1 \sim 0.3 \tag{1.67}$$

for the cosmological constant.

While the energy per unit comoving volume stored in pressureless matter is conserved, this is not the case of radiation ($\propto a^{-1}$ owing to the redshift of photons' wavelengths) or a cosmological constant ($\propto a^3$). This follows from the fact that time-translational is broken (In classical physics, invariance under time-translations implies a conserved charge from Noether's theorem: Energy). Is this a problem ? No, as there is no global energy conservation in GR. Namely, $\nabla_\mu T^{\mu\nu} = 0$ is a local conservation law, which cannot be brought to an integral form due to the presence of ∇_μ rather than ∂_μ.

We can now formally integrate the first Friedmann equation to find $t(a)$ or, equivalently, $a(t)$ for a given set of constituents. Namely, the first Friedmann equation can now be written as

$$H^2 = H_0^2 \left(\sum_X \Omega_X a^{-3(1+w_X)} + \Omega_K a^{-2}\right) \,, \tag{1.68}$$

so that $t(a)$ can be expressed as the integral

$$t = \int \frac{da}{\dot{a}} = \int \frac{da}{aH} = \int \frac{da}{aH_0\sqrt{\sum_X \Omega_X a^{-3(1+w_X)} + \Omega_K a^{-2}}} \,. \tag{1.69}$$

Here, w_X is the equation of state parameter of the component X and it is understood that $\Omega_K = -K/H_0^2$. This integral is generally difficult to evaluate unless one of the components dominates over the others.

Exercices

[Universe with matter and curvature]

We wish to obtain the time dependence of a in *dust-dominated* universe (matter+curvature). For this purpose, we momentarily give up our normalization $a_0 = 1$ so as to choose $k = 0, \pm 1$.

(a) Show that the first Friedmann equation becomes

$$\dot{a}^2 = \frac{c_m}{a} - k \ . \tag{1.70}$$

where $c_m = \Omega_m H_0^2 a_0^3$.

(b) Show that, for a flat universe with $k = 0$, the solution is

$$a(t) \propto t^{2/3} \propto \eta^2 \ . \tag{1.71}$$

This solution is known as *Einstein-de Sitter (EdS) spacetime*.

(c) For $k = +1$, we have a re-collapsing universe with $a_{\text{max}} = c_m$. The equation can be solved in closed form for $t(a)$. The curve $a(t)$ is a cycloid, as is best seen upon going to conformal time η and writing the solution in parametrized form. Namely, write

$$\frac{dt}{a(t)} = + \frac{da}{\sqrt{c_m a - a^2}} \ , \tag{1.72}$$

and show that it integrates to

$$\eta = 2 \arctan\left(\sqrt{\frac{a}{c_m - a}} \right) \ . \tag{1.73}$$

Using trigonometric identities such as $\sin^2(x/2) = \tan^2(x/2)/(1+\tan^2(x/2)) = (1-\cos x)/2$, demonstrate that

$$a(\eta) = \frac{a_{\text{max}}}{2}(1 - \cos\eta) \ , \qquad t(\eta) = \frac{a_{\text{max}}}{2}(\eta - \sin\eta) \ . \tag{1.74}$$

The maximum radius is reached when $\eta = \pi$, i.e. $t_{\text{max}} = (\pi/2)a_{\text{max}}$.

Analogously, for $k = -1$, the first Friedmann equation can be solved in parametric form upon replacing the trigonometric functions by hyperbolic functions. As we shall see below, open universes containing only non-relativistic matter are one historically important example of FRW universes with non-negligible curvature.

[Distances and the angular size of CMB fluctuations]

The comoving distance D_C between earth at $\mathcal{O}$ and a distant object at $\mathcal{O}'$ is the radial coordinate χ, which is fixed by the requirement that a photon emitted from $\mathcal{O}'$ at scale factor a reaches $\mathcal{O}$ at present time $a = a_0 = 1$. Hence, integrating the infinitesimal distance $c\Delta t$ travelled by a photon divided the scale factor at that time, we have

$$\chi(t) = \int_t^{t_0} \frac{dt}{a(t)} = \int_a^1 \frac{1}{a\dot{a}} da = \int_a^1 \frac{da}{a^2 H} = \eta_0 - \eta \ , \tag{1.75}$$

where η_0 is the present-day value of the conformal time. D_C is the fundamental distance measure in large scale structure.

(d) Consider a photon emitted at a radial coordinate cH_0^{-1} from Earth at time $t = 0$. Assuming a EdS Universe, calculate both the comoving separation $\chi(t) \equiv D_C(t)$ and the physical separation $r_{\text{phys}}(t)$ between the photon and Earth as a function of time t. Draw (approximately) $r_{\text{phys}}(t)$ as a function of t/t_0, where t_0 is the present age of the Universe and compare its behaviour to that of $\chi(t)$.

Another useful distance is the *comoving angular diameter distance* D_{CA}, which relies on the comoving (rather than the physical) size of an object. Namely, if R is the physical extent of an object at a certain radial comoving distance χ, the comoving angular diameter distance inferred by an observer at the origin of coordinates is

$$D_{CA} = \frac{R/a}{\theta} = \frac{\sqrt{g_{\theta\theta}}\theta/a}{\theta} = \sqrt{\gamma_{\theta\theta}} = \sin_K \chi \ , \tag{1.76}$$

where we have assumed that the object has constant longitude ϕ.

In the early 1990s, observations indeed appeared to favor a low density universe with $\Omega_m \sim 0.2$ and thus prompted great interest in the cosmology of open universes. Measurements of the comoving angular diameter distance D_{CA} from CMB anisotropies eventually led to the abandonment of the $\Omega_m = 0.2$ open model in favor of a flat, dark energy dominated universe.

(e) To see this, we will calculate D_{CA} to the CMB last scattering surface in an open, matter-dominated universe. The conformal time is

$$\eta = \int \frac{da}{a^2 H_0 \sqrt{\Omega_m a^{-3} + \Omega_K a^{-2}}} = \frac{2}{H_0\sqrt{\Omega_K}} \operatorname{arctanh} \sqrt{\frac{\Omega_K a}{\Omega_K a + \Omega_m}} \ , \tag{1.77}$$

where we have taken $\eta = 0$ at $a = 0$. Now, the CMB formed very early in the universe, $a \ll 1$, so the CMB photons we observe today came to us from a radial comoving distance of $\chi \approx \eta_0 = \eta(a = 1)$. Show that the comoving angular diameter distance to the CMB is

$$D_{CA} = \frac{1}{H_0\sqrt{\Omega_K}} \sinh\left(2\operatorname{arctanh}\sqrt{\Omega_K}\right) = \frac{2}{H_0 \Omega_m} \tag{1.78}$$

in such a universe.

(f) We will see later that the comoving sound horizon at the time of (CMB) decoupling is $r_s \approx 105\ h^{-1}\text{Mpc}$. Calculate the angular size $\theta(r_s)$ (in deg) under which this comoving scale is seen from earth.

(g) CMB experiments show that $\theta(r_s) \sim 1$ deg. What are the implications of this measurement for open models with low matter content ?

Chapter 2

Linear cosmological perturbation theory

We introduce scalar perturbations and discuss the physics of the cosmic microwave background (CMB) and its angular power spectrum. We work at linear order in perturbations.

2.1 Perturbed metric

With the metric signature $(-,+,+,+)$ and the comoving coordinates $(\eta, \mathbf{x})$ (η is the conformal time), the most general perturbed FRW metric can be written as

$$ds^2 = a^2\Big\{ -\big(1+2A\big)d\eta^2 - 2B_i d\eta dx^i + \Big[(1-2H_L)\gamma_{ij} + H_{ij}\Big]dx^i dx^j\Big\} \,, \tag{2.1}$$

where H_{ij} is a traceless tensor, i.e. $\gamma^{ij}H_{ij} = 0$ (its trace is absorbed in H_L). At linear order in perturbations,

- You can treat the perturbation variables as 3-tensors and raise/lower their indices with γ_{ij}.
- You must raise/lower components of 4-vectors with $g_{\mu\nu}$.

A, B_i, H_L and H_{ij} are all functions of time and space and includes *scalar, vector and tensor type perturbations*. In what follows, we will assume a flat FLRW background in all illustrations, so that $\gamma_{ij} = \delta_{ij}$ (though I'll keep the notation γ_{ij}).

2.1.1 Scalar-Vector-Tensor (SVT) decomposition

While the lapse function A and the curvature perturbation H_L are 2 scalar d.o.f. (degrees of freedom) and, as such, cannot be simplified, any vector field can be decomposed into a *gradient* and a *rotation*.

Analogously, it is convenient to decompose spatial vectors like B_i into their *longitudinal* (or irrotational or curl-free) and *transverse* (or solenoidal or divergence-free) piece:

$$B_i = B_i^{\|} + B_i^{\perp} \equiv \partial_i b + B_i^{\perp} \,, \tag{2.2}$$

where

$$\text{longitudinal}: \quad \epsilon^{ijk}\partial_j B_k^{\|} \equiv \vec{\nabla}\times\vec{B} = 0 \tag{2.3}$$

$$\text{transverse}: \quad \partial^i B_i^{\perp} = \vec{\nabla}\cdot\vec{B} = 0\ , \tag{2.4}$$

where ϵ_{ijk} is the Levi-Civita tensor. The transverse part $B_i^{\perp}$ can be further decomposed into poloidal and toroidal components. The potential b is a spin 0 perturbation, whereas $B_i^{\perp}$ is a spin 1 perturbation.

Similarly, traceless spatial tensors like H_{ij} can be decomposed into

$$H_{ij} = H_{ij}^{\top} + H_{ij}^{\|} + H_{ij}^{\perp}\ , \tag{2.5}$$

where $H_{ij}^{\top}$ is transverse, while the divergences of $H_{ij}^{\|}$ and $H_{ij}^{\perp}$ are longitudinal and transverse vectors, respectively. Namely,

$$\text{transverse}: \quad \gamma^{ij}H_{ij}^{\top} = 0 \quad \text{and} \quad \partial^i H_{ij}^{\top} = 0 \tag{2.6}$$

$$\text{longitudinal divergence}: \quad \gamma^{ij}H_{ij}^{\|} = 0 \quad \text{and} \quad \epsilon^{ijk}\partial_j(\partial^l H_{kl}^{\|}) = 0 \tag{2.7}$$

$$\text{transverse divergence}: \quad \gamma^{ij}H_{ij}^{\perp} = 0 \quad \text{and} \quad \partial^i(\partial^j H_{ij}^{\perp}) = 0\ . \tag{2.8}$$

$H_{ij}^{\|}$ is a scalar perturbation, whereas $H_{ij}^{\perp}$ is a vector perturbation. The transverse piece $H_{ij}^{\top}$ corresponds to a genuine tensor perturbation. The 2 independent d.o.f. are the polarizations of gravity waves (a massless spin-2 field).

2.1.2 Gauge transformations

In General Relativity, since one can perform arbitrary coordinate transformations, the notion of perturbation is ambiguous because there is no unique decomposition of a variable, say the density $\rho(t,\mathbf{x})$, into a background or homogeneous piece $\bar{\rho}(t)$ and a perturbation $\delta\rho(t,\mathbf{x})$. Spatial averaging depends on the choice of the hypersurfaces of constant time and is practically impossible on superhorizon scales (in addition to the fact that it does not commute with derivatives).

- Consider a perturbation of the form $\rho(t,\mathbf{x}) = kt + \epsilon\sin x$, $\epsilon \ll 1$. Averaging over x at a given time t, we infer a background $\bar{\rho}(t) = kt$ and a perturbation $\delta\rho(t,\mathbf{x}) = \epsilon\sin x$. However, changing coordinates to (t',x') such that $x = x'$ and the new time coordinate is $t' = t + (\epsilon/k)\sin x$, we find that, in the new coordinates, the background density is $\bar{\rho}'(t') = kt'$ while the perturbation vanishes, $\delta\rho'(t',x') = 0$.

- To define perturbations, one must compare the actual perturbed spacetime $\mathcal{M}$ to a homogeneous and isotropic FLRW reference spacetime $\overline{\mathcal{M}}$ (endowed with a FLRW metric). Once a coordinate system - or a chart U - is set, the density perturbation at the point $X \in \mathcal{M}$ is

$$\begin{aligned}\delta\rho(x^\mu) &= \rho(x^\mu)\big|_{\mathcal{M}} - \bar{\rho}(x^\mu)\big|_{\overline{\mathcal{M}}} \\ &\equiv \rho(x^\mu) - \bar{\rho}(x^\mu)\ .\end{aligned} \tag{2.9}$$

- A gauge choice corresponds to a specific map or diffeomorphism between $\mathcal{M}$ and $\overline{\mathcal{M}}$. This defines a specific *slicing* (hypersurfaces of constant time t) and *threading* (timelike wordlines of constant $\mathbf{x}$)) of spacetime. The example above shows that one can for instance choose the slicing such that it coincides with hypersurfaces of constant energy density.

- A gauge transformation amounts to changing the map between $\mathcal{M}$ and $\overline{\mathcal{M}}$ *without changing the chart* U. This induces a coordinate change on $\mathcal{M}$ known as a *gauge transformation*, which can be viewed as an infinitesimal coordinate change (although it is technically slighty different). Under such a transformation, the coordinates of the point $X \in \mathcal{M}$ change to $x^{'\mu} = x^\mu + \epsilon^\mu$. Because perturbations are small in the Early Universe, one treats ϵ^μ as a small quantity.

- For our density ρ, the density perturbation after a gauge transformation becomes

$$\begin{aligned} \delta\rho'(x^{'\mu}) &= \rho'(x^{'\mu}) - \bar{\rho}'(x^{'\mu}) \\ &= \rho(x^\mu) - \bar{\rho}'(x^{'\mu}) && (\rho \text{ is a scalar}) \\ &= \rho(x^\mu) - \bar{\rho}(x^{'\mu}) && (\text{same solution on } \overline{\mathcal{M}}) \\ &= \rho(x^{'\mu} - \epsilon^\mu) - \bar{\rho}(x^{'\mu}) \\ &= \rho(x^{'\mu}) - \epsilon^\mu \partial_\mu \rho(x^{'\mu}) - \bar{\rho}(x^{'\mu}) \ . \end{aligned} \tag{2.10}$$

  Therefore,

$$\delta\rho' = \delta\rho - \epsilon^\mu \partial_\mu \rho = \delta\rho - \epsilon^0 \partial_0 \bar{\rho} \ . \tag{2.11}$$

- A gauge transformation involves $\epsilon^\mu = (\epsilon^0, \epsilon^i)$, which can be decomposed into two spin 0 perturbations ϵ_0 and $\epsilon_i^{\parallel}$, and one spin 1 perturbation $\epsilon_i^{\perp}$.

- The metric Eq.(2.1) has 10 dynamical d.o.f. but, since a gauge transformation $x^\mu \to x^{\mu'}$ involves 4 functions ϵ^μ, only $10 - 4 = 6$ of the metric d.o.f. are physical. Furthermore, the 6 physical d.o.f. decompose into two spin 0, one spin 1 and one spin 2 perturbations.

- Since gauge transformations do not involves spin 2 quantities, $H_{ij}^{\top}$ is gauge invariant, i.e. it remains the same under any redefinition of the background. By contrast, all the other variables are not invariant under gauge transformations.

A gauge corresponds to a specific parametrization of the 6 physical d.o.f. of the perturbed metric. Two popular choices are the *synchronous gauge*, which corresponds to $A = B_i = 0$, and the *conformal Newtonian* or *longitudinal* gauge, in which the 2 scalars d.o.f. are $A \equiv \Psi$ and $H_L = \Phi$.

In what follows, we shall consider only scalar degrees of freedom and work in the (conformal) Newtonian gauge. Its name arises from the fact that, in the non-relativistic limit, the perturbed 00 Einstein equation takes the form of Poisson equation with a Newtonian gravitational potential Ψ. The corresponding metric is

$$ds^2 = a^2 \Big[-\big(1 + 2\Psi\big) d\eta^2 + \big(1 - 2\Phi\big)\gamma_{ij} dx^i dx^j \Big] \ , \tag{2.12}$$

where Φ is a perturbation to the curvature of constant time hypersurfaces. With our sign convention, an overdense region has $\Psi < 0$ and $\Phi < 0$. Furthermore, $\Psi = \Phi$ in the absence of anisotropic stress.

In the early Universe, Ψ and Φ are small perturbations – $|\Psi| \sim |\Phi| \sim 10^{-5}$ – and similarly for perturbations to the stress-energy tensor $T^{\mu\nu}$. This justifies our linear perturbation analysis.

2.2 Boltzmann equations for photons and baryons

Although CDM makes the majority of non-relativistic matter in the Universe, we shall ignore it for simplicity and consider a two-component photon-baryon fluid. Photons and baryons are both affected by gravity and are coupled by Compton scattering (i.e. scattering of photons off free electrons). Therefore, we need to simultaneously solve for both components. This can be carried out in a systematic way by writing down a Boltzmann equation for each species in the Universe. We shall focus on the photon distribution, while we only briefly discuss baryons.

Here, we are interested in more detailed information, not just the number density of a given species which is the integral of the distribution function over all momenta. Schematically, the Boltzmann equation takes the form

$$\frac{df}{d\eta} = \left(\frac{df}{d\eta}\right)_{\rm coll} , \tag{2.13}$$

where the right-hand side contains all possible collision terms. In the absence of collisions, $df/d\eta = 0$. The difficulty is that phase space elements are moving in time in a complicated way owing to the nontrivial metric Eq.(2.12).

The distribution function $f = f(\eta, x^i, p_\mu)$ generally depends on conformal time η, position x^i and (covariant) 4-momentum p_μ, such that the on-shell condition $g_{\mu\nu}p^\mu p^\nu = -m^2$ is satisfied. Therefore, f can also be written as a function $f(\eta, x^i, p, \hat{n}^i)$. Here, $p = \sqrt{g_{ij}p^ip^j}$ is the modulus of the *physical* 3-momentum p^i and $\hat{n}^i$ are the components of the unit vector $\hat{\mathbf{n}}$ which indicates the direction of propagation of the particle. By definition, we have

$$\hat{\mathbf{n}} \cdot \hat{\mathbf{n}} = \gamma_{ij}\hat{n}^i\hat{n}^j = 1 . \tag{2.14}$$

For massive particles, $p^i = m dx^i/d\tau$ where τ is the proper time of the particle whereas, for massless particles like photons, $p^i = dx^i/d\lambda$ where λ parametrizes the particle's path. Note that $p^i \neq p\hat{n}^i$. Therefore, the total time derivative in the left-hand side of Eq.(2.13) can be written as

$$\frac{df}{d\eta} = \frac{\partial f}{\partial \eta} + \frac{\partial f}{\partial x^i}\frac{dx^i}{d\eta} + \frac{\partial f}{\partial p}\frac{dp}{d\eta} + \frac{\partial f}{\partial \hat{n}^i}\frac{d\hat{n}^i}{d\eta} . \tag{2.15}$$

The last term does not contribute at first order in perturbations because i) the unperturbed distribution $\bar{f}$ is the simple Bose-Einstein function which depends only on p, not $\hat{n}^i$; and ii) $d\hat{n}^i/d\eta$ is non-zero only in the presence of metric perturbations. We must now evaluate $dx^i/d\eta$ and $dp/d\eta$ along the particle trajectories.

2.2.1 Propagation in a perturbed FRW universe

Let us consider first massless particles. A few useful relations:

- The on-shell condition $g_{\mu\nu}p^\mu p^\nu = 0$ generalizes the special relativistic expression $E^2 - p^2 = 0$ as follows:

$$a^2\left(1+2\Psi\right)\left(p^0\right)^2 = g_{ij}p^ip^j = p^2 , \tag{2.16}$$

which implies

$$p^0 = \frac{p}{a}\left(1-\Psi\right) . \tag{2.17}$$

Here, $E = ap^0 \propto a^{-1}$ is the energy that would be measured by a comoving observer in an unperturbed universe (such an observer would have a 4-velocity $u^\mu = a^{-1}(1,0,0,0)$ in our coordinate system). This tells us that p^0 decreases as photons move out of a potential well (with $\Psi < 0$). This is nothing but the redshift of photons in a gravitational field.

- Writing $p^i = C\hat{n}^i$, we determine C from $p^2 = g_{ij}p^ip^j$. We find

$$C = \frac{p}{a}\big(1 + \Phi\big) \ , \tag{2.18}$$

which implies

$$\frac{dx^i}{d\eta} = \frac{p^i}{p^0} = \hat{n}^i\big(1 + \Psi + \Phi\big) \ . \tag{2.19}$$

In other words, the coordinate velocity of the photons slows down when they travel through an overdense region.

- To compute $dp/d\eta$, start from the geodesic equation. At first order in perturbations, we find

$$\frac{1}{p}\frac{dp}{d\eta} = -\mathcal{H} + \dot{\Phi} - \hat{n}^i\partial_i\Psi \ , \tag{2.20}$$

where an overdote denotes a derivative w.r.t. conformal time and $\mathcal{H} = \dot{a}/a$ is the conformal Hubble rate. This equation describes the change in the photon momenum as it moves through a perturbed FRW universe. The first term accounts for the loss of momentum due to the expansion of the universe. The second says that a photon in a deepening gravitational well ($\dot{\Phi} < 0$) loses energy (i.e. the magnitude of its redshift increases). The third term represents an energy gain when the photon travels towards the center of a potential well ($\hat{n}^i\partial_i\Psi < 0$).

Similar relations can be derived for massive particles such as baryons using the same equations. We find

$$\frac{dx^i}{d\eta} = \left(\frac{p}{E}\right)\hat{n}^i\big(1 + \Psi + \Phi\big) \tag{2.21}$$

$$\frac{1}{p}\frac{dp}{d\eta} = -\mathcal{H} + \dot{\Phi} - \left(\frac{E}{p}\right)\hat{n}^i\partial_i\Psi \ , \tag{2.22}$$

where $E = \sqrt{p^2 + m^2}$. Note that, in the non-relativistic limit $p \ll m$, $dx^i/d\eta \to (p/m)\hat{n}^i$ and $dp/d\eta \to -m\hat{n}^i\partial_i\Psi$ and we recover the usual non-relativistic Boltzmann equation.

2.2.2 The collision term: Compton scattering

The scattering process of interest is

$$e^-(\mathbf{k}) + \gamma(\mathbf{p}) \rightleftharpoons e^-(\mathbf{k}') + \gamma(\mathbf{p}') \ . \tag{2.23}$$

For photons, the collision term is schematically given by

$$\left(\frac{df_\gamma}{d\eta}\right)_{\text{coll}} = \sum_{\mathbf{k},\mathbf{k}',\mathbf{p}'} \big|\mathcal{M}\big|^2\big[f_e(\mathbf{k}')f_\gamma(\mathbf{p}') - f_e(\mathbf{k})f_\gamma(\mathbf{p})\big] \ , \tag{2.24}$$

where f_e and f_γ are the (kinetic equilibrium) distribution functions of the free electrons and photons, respectively, and $\mathcal{M}$ is the scattering amplitude. We have neglected stimulated emission and Pauli blocking, which would lead to factors of $1 + f_\gamma$ and $1 - f_e$. At first order, this turns out to be a valid assumption.

In the non-relativistic limit $T \ll m_e$, Compton scattering is nearly elastic, $p \simeq p'$, and the fractional energy change in a single Compton collision is very small (of order of the baryon velocity $\mathbf{v}_b$). This limit is known as Thompson scattering (no energy exchange, only scattering of momenta). In a first approximation, Feynman rules show that the amplitude $\mathcal{M}$ is constant:

$$\left|\mathcal{M}\right|^2 = 8\pi\sigma_T m_e^2 \ , \tag{2.25}$$

where $\sigma_T = 6.65 \times 10^{-25}$ cm^2 is the Thompson cross-section. However, this ignores i) the angular dependence $\propto 1+\cos^2(\mathbf{p}\cdot\mathbf{p}')$ and ii) the polarization dependence $\propto |\epsilon_\mathbf{p} \cdot \epsilon_{\mathbf{p}'}|^2$. The dependence on polarization means that a small level of the CMB will be polarized owing to Compton scattering. We shall not discuss this here.

To proceed further, we expand the photon distribution function f_γ around its zero-order Planckian value (the spin degeneracy factor $g_\gamma = 2$ is not included)

$$\bar{f}_\gamma(\eta, p) = \frac{1}{e^{p/\bar{T}} - 1} \ , \tag{2.26}$$

where, for simplicity, we have ignored any non-zero chemical potential and the zeroth order temperature satisfies $\bar{T}(\eta) \propto a(\eta)^{-1}$. We have

$$f_\gamma(\eta, x^i, p, \hat{n}^i) = \left\{ \exp\left[\frac{p}{\bar{T}(\eta)\big(1 + \Theta(\eta, x^i, \hat{n}^i)\big)}\right] - 1 \right\}^{-1} \ , \tag{2.27}$$

where $\Theta(\eta, x^i, \hat{n}^i)$ is the relative temperature perturbation. Note that it depends on both x^i (there are inhomogeneities) and $\hat{n}^i$ (there are anisotropies) but not on p (because photons are all affected in the same way regardless of their frequency). At first order, we have

$$f_\gamma \approx \bar{f}_\gamma - p\frac{\partial \bar{f}_\gamma}{\partial p}\Theta = \bar{f}_\gamma + \delta f_\gamma \ . \tag{2.28}$$

After a lengthy but straightforward calculation, the photon collision term can eventually be written as

$$\left(\frac{df_\gamma}{d\eta}\right)_{\rm coll} = -p\frac{\partial \bar{f}_\gamma}{\partial p}\dot{\tau}_c\Big[\Theta_0(\eta, x^i) - \Theta(\eta, x^i, \hat{n}^i) + \hat{\mathbf{n}} \cdot \mathbf{v}_b\Big] \tag{2.29}$$

where

$$\dot{\tau}_c = \frac{d\tau_c}{d\eta} = \sigma_T n_e a \tag{2.30}$$

is the Thompson optical depth per unit (conformal) time, and

$$\Theta_0(\eta, x^i) \equiv \frac{1}{4\pi}\int d\Omega\, \Theta(\eta, x^i, \hat{n}^i) \tag{2.31}$$

is the *monopole* part of the temperature fluctuation and, thus, does not depend on the direction vector $\hat{\mathbf{n}}$. Θ_0 represents the deviation of the temperature monopole at a given point in space from its average over all space. We will later generalize Eq.(2.31) to higher order multipoles of the photon temperature fluctuation. Note that, since $\rho_\gamma \propto T^4$,

$$\Theta_0 = \frac{1}{4}\frac{\delta\rho_\gamma}{\bar{\rho}_\gamma} = \frac{1}{4}\delta_\gamma \ , \tag{2.32}$$

where δ_γ is the fractional perturbation in the photon energy density.

The first term in the right-hand side of Eq.(2.29) is the scattering contribution into the beam, the second is the removal of photons from the beam due to Thompson scattering, and the third is a Doppler effect, which arises from the fact that electrons are not at rest relative to the CMB (cosmic) frame.

2.2.3 Boltzmann hierarchy for photons

We now combine Eqs.(2.15), (2.19), (2.20) and (2.29) and write the photon Boltzmann equation as

$$\frac{\partial f_\gamma}{\partial \eta} + \hat{n}^i \frac{\partial f_\gamma}{\partial x^i} - \mathcal{H}p\frac{\partial f_\gamma}{\partial p} + p\frac{\partial \bar{f}_\gamma}{\partial p}\big(\dot{\Phi} - \hat{n}^i\partial_i\Psi\big) = -p\frac{\partial \bar{f}_\gamma}{\partial p}\dot{\tau}_c\Big[\Theta_0 - \Theta + \hat{\mathbf{n}}\cdot\mathbf{v}_b\Big] \tag{2.33}$$

where we have retained zeroth and first order terms only. In the left-hand side, the first two terms are familiar from the Newtonian limit, the third describes energy loss due to the expansion of the Universe, and the last two encode the effects of over-/underdense regions on the photon distribution.

- Exploit the fact that the unperturbed photon distribution satisfies

  $$\frac{d\bar{f}_\gamma}{d\eta} = \frac{\partial \bar{f}_\gamma}{\partial \eta} - \mathcal{H}p\frac{\partial \bar{f}_\gamma}{\partial p} = 0 \ , \tag{2.34}$$

  which represents the free-streaming of photons in an expanding universe (the collision term is first order in perturbations). As a result, the average photon distribution function does not change its shape: the photon blackbody spectrum remains blackbody as the Universe expands.

- Use the relations

  $$\frac{\partial}{\partial \eta}\left[-p\frac{\partial \bar{f}_\gamma}{\partial p}\Theta\right] = -p\frac{\partial}{\partial p}\left(\frac{\partial \bar{f}_\gamma}{\partial \bar{T}}\right)\dot{\bar{T}}\Theta - p\frac{\partial \bar{f}_\gamma}{\partial p}\dot{\Theta} \tag{2.35}$$

  $$= -\mathcal{H}p\frac{\partial}{\partial p}\left(p\frac{\partial \bar{f}_\gamma}{\partial p}\right)\Theta - p\frac{\partial \bar{f}_\gamma}{\partial p}\dot{\Theta}$$

  $$-\mathcal{H}p\frac{\partial}{\partial p}\left[-p\frac{\partial \bar{f}_\gamma}{\partial p}\Theta\right] = \mathcal{H}p\frac{\partial \bar{f}_\gamma}{\partial p}\Theta + \mathcal{H}p^2\frac{\partial^2 \bar{f}_\gamma}{\partial p^2}\Theta \ , \tag{2.36}$$

  the first of which follows from $\partial \bar{f}_\gamma/\partial \bar{T} = -(p/\bar{T})\partial \bar{f}_\gamma/\partial p$,

We eventually arrive at

$$\dot{\Theta} + \hat{n}^i\partial_i\Theta = -\hat{n}^i\partial_i\Psi + \dot{\Phi} + \dot{\tau}_c\Big[\Theta_0 - \Theta + \hat{\mathbf{n}}\cdot\mathbf{v}_b\Big] \ . \tag{2.37}$$

This differential equation coupling Θ, Ψ, Φ and $\mathbf{v}_b$ together is transformed further as follows:

- When the perturbations are small, as is the case until decoupling, there is an added benefit of Fourier transforming Eq.(2.37) because the different Fourier modes evolve independently. For the potential Ψ and Φ, the Fourier transform is immediate:

$$\Psi(\eta, \mathbf{x}) = \int \frac{d^3k}{(2\pi)^3}\, \Psi(\eta, \mathbf{k})\, e^{i\mathbf{k}\cdot\mathbf{x}} \tag{2.38}$$

$$\Phi(\eta, \mathbf{x}) = \int \frac{d^3k}{(2\pi)^3}\, \Phi(\eta, \mathbf{k})\, e^{i\mathbf{k}\cdot\mathbf{x}} \,. \tag{2.39}$$

As discussed above, the baryon velocity field $\mathbf{v}_b(\eta, \mathbf{x})$ can generally be written as the sum of a longitudinal and transverse piece $v_b^{\parallel}(\eta, \mathbf{x})$ and $v_b^{\perp}(\eta, \mathbf{x})$, the first of which is the scalar perturbation. It is convenient to write it as the gradient of a potential:

$$\begin{aligned} v_b^{\parallel}(\eta, \mathbf{x}) &= \boldsymbol{\nabla}_{\mathbf{x}} \int \frac{d^3k}{(2\pi)^3} \left(-\frac{V_b(\mathbf{k})}{k} \right) e^{i\mathbf{k}\cdot\mathbf{x}} \\ &= \int \frac{d^3k}{(2\pi)^3} \Big(-i\hat{\mathbf{k}} V_b(\mathbf{k}) \Big) e^{i\mathbf{k}\cdot\mathbf{x}} \,. \end{aligned} \tag{2.40}$$

This definition ensures that the velocity potential has dimensions of velocity (i.e. is dimensionless in natural units). The minus sign is added for future convenience. Finally, for the temperature fluctuation $\Theta(\eta, \mathbf{x}, \hat{\mathbf{n}})$, we write

$$\Theta(\eta, \mathbf{x}, \hat{\mathbf{n}}) = \int \frac{d^3k}{(2\pi)^3} \Theta(\eta, \mathbf{k}, \hat{\mathbf{n}}) e^{i\mathbf{k}\cdot\mathbf{x}} \,. \tag{2.41}$$

Defining $\mu = \hat{\mathbf{k}} \cdot \hat{\mathbf{n}}$ as the cosine of the angle between the propagation direction of the plane-wave perturbation and the photon direction, this implies

$$\hat{n}^i \partial_i \Theta(\eta, \mathbf{x}, \hat{\mathbf{n}}) = \int \frac{d^3k}{(2\pi)^3}\, ik\mu\, \Theta(\eta, \mathbf{k}, \hat{\mathbf{n}})\, e^{i\mathbf{k}\cdot\mathbf{x}} \,. \tag{2.42}$$

With aid of these results, we can Fourier transform Eq.(2.37) to obtain

$$\dot{\Theta} + ik\mu\Theta = -ik\mu\Psi + \dot{\Phi} + \dot{\tau}_c \Big[\Theta_0 - \Theta - i\mu V_b \Big] \tag{2.43}$$

where it is understood that the potential Ψ, curvature Φ and baryon velocity V_b are function of $(\eta, \mathbf{k})$, whereas $\Theta = \Theta(\eta, \mathbf{k}, \hat{\mathbf{n}})$.

- It is convenient to decompose the temperature fluctuation $\Theta(\eta, \mathbf{k}, \hat{\mathbf{n}})$ into its multipoles $\Theta_l(\eta, \mathbf{k})$ defined through the relation

$$\Theta(\eta, \mathbf{k}, \hat{\mathbf{n}}) = \sum_l (-i)^l \Theta_l(\eta, \mathbf{k}) \mathcal{P}_l(\hat{\mathbf{k}} \cdot \hat{\mathbf{n}}) \,, \tag{2.44}$$

where $\mathcal{P}_l(\mu)$ are Legendre polynomials, with $\mathcal{P}_0 = 1$, $\mathcal{P}_1 = \mu$ and $\mathcal{P}_2 = (3\mu^2 - 1)/2$ etc. The multipole moments of the photon temperature field are thus given by

$$\Theta_\ell(\eta, \mathbf{k}) = \big(2\ell + 1\big) \frac{i^\ell}{2} \int_{-1}^{+1} d\mu\, \mathcal{P}_l(\mu) \Theta(\eta, \mathbf{k}, \hat{\mathbf{n}}) \,. \tag{2.45}$$

In particular, the monopole, dipole and quadrupole read

$$\Theta_0(\eta, \mathbf{k}) = \frac{1}{2}\int_{-1}^{+1} d\mu\, \Theta(\eta, \mathbf{k}, \hat{\mathbf{n}}) \tag{2.46}$$
$$\Theta_1(\eta, \mathbf{k}) = \frac{3i}{2}\int_{-1}^{+1} d\mu\, \mu\, \Theta(\eta, \mathbf{k}, \hat{\mathbf{n}}) \tag{2.47}$$
$$\Theta_2(\eta, \mathbf{k}) = -\frac{5}{4}\int_{-1}^{+1} d\mu\, \left(3\mu^2 - 1\right) \Theta(\eta, \mathbf{k}, \hat{\mathbf{n}}) \; . \tag{2.48}$$

As a result, Eq.(2.37) can be broken up into a hierarchy of coupled differential equation: the *Boltzmann hierarchy*. After some manipulations (left as an exercise), we obtain the following evolution equations for the first two moments:

$$\dot{\Theta}_0 = -\frac{k}{3}\Theta_1 + \dot{\Phi} \tag{2.49}$$
$$\dot{\Theta}_1 = k\left(\Theta_0 + \Psi - \frac{2}{5}\Theta_2\right) - \dot{\tau}_c\big(\Theta_1 - V_b\big) \; . \tag{2.50}$$

The system is not closed as it involves Θ_2.

- The temperature quadrupole Θ_2 is related to the photon anisotropic stress π_γ through

$$\pi_\gamma = \frac{12}{5}\Theta_2 \; . \tag{2.51}$$

The photon anisotropic stress represents the viscosity that arises from the diffusion of "hotter" photons into regions with lower temperature.

2.2.4 Energy and momentum conservation for baryons

Boltzmann equations can also be derived for the electrons and protons. Both species are coupled by Coulomb scattering, whose rate is much larger than the expansion rate at all epochs of interest. This tight coupling forces the electron and proton overdensities to a common baryon overdensity δ_b (monopole) and velocity $\mathbf{v}_b$ (dipole). Higher multipoles (quadrupole etc.) can be safely neglected. The proton-photon scattering can be ignored (it is inversely proportional to the proton mass $m_p \gg m_e$) so that the interaction between baryons and electrons is driven by the Compton scattering of electrons. We shall not derive the Boltzmann equation for f_e and f_p as it closely follows the procedure outlined above.

Integrating the electron and proton Boltzmann equations (i.e. taking the zeroth moment) and adding them together yields an energy conservation law for baryons:

$$\dot{\delta}_b = -kV_b + 3\dot{\Phi} \; . \tag{2.52}$$

Similarly, taking the first moment furnishes another conservation law: the conservation of baryon momentum,

$$\dot{V}_b = -\mathcal{H}V_b + k\Psi + \dot{\tau}_c\frac{\big(\Theta_1 - V_b\big)}{R} \; , \tag{2.53}$$

where $R = 3\bar{\rho}_b/4\bar{\rho}_\gamma$ is the ratio of baryon to photon energy densities. Note the appearance of a factor of $1/\bar{\rho}_b$ which, although photons scatter primarily off electrons, arises from the fact that electrons are tightly coupled to protons via Coulomb scattering. This remains valid even in the presence of a small fraction of neutral hydrogen, and helium atoms or ions. If the protons were infinitely heavy, i.e. $\rho_b \to \infty$, Compton scattering would not change the electron velocity at all.

To conclude, let us mention that, had we considered CDM as well, we should also have added the following equations:

$$\begin{aligned} \dot{\delta}_c &= -kV_c + 3\dot{\Phi} \\ \dot{V}_c &= -\mathcal{H}V_c + k\Psi \ . \end{aligned} \tag{2.54}$$

where V_c is the velocity of the dark matter.

2.3 Baryon-photon fluid

Eqs.(2.49), (2.50), (2.52) and (2.53) are 4 differential equations for the 7 functions Θ_0, Θ_1 Θ_2, Ψ, Φ, δ_b and V_b. Therefore, they must supplemented by additional equations (e.g. Einstein equations etc.). Nonetheless, we will see that it is possible to obtain solutions for a linear combination of Θ_0 and Ψ in the so-called tight coupling limit.

Exercices

[Tight-coupling approximation]

The first two moments of the photon and baryon Boltzmann equations read

$$\dot{\Theta}_0 = -\frac{k}{3}\Theta_1 + \dot{\Phi} \tag{2.55}$$

$$\dot{\Theta}_1 = k\left(\Theta_0 + \Psi - \frac{2}{5}\Theta_2\right) - \dot{\tau}_c\big(\Theta_1 - V_b\big) \tag{2.56}$$

$$\dot{\delta}_b = -kV_b + 3\dot{\Phi} \tag{2.57}$$

$$\dot{V}_b = -\mathcal{H}V_b + k\Psi + \dot{\tau}_c\frac{\big(\Theta_1 - V_b\big)}{R}\ . \tag{2.58}$$

V_b represents the baryon bulk velocity. Θ_1 should be interpreted as the bulk velocity of the *photon temperature perturbation*.

This system of coupled ODEs simplifies in the so-called *tight-coupling limit*. Let $k \sim L^{-1}$ be the wavenumber of the fluctuations. There are two important characteristic timescales:

$$k^{-1} \quad = \quad \text{travel time across the perturbation} \tag{2.59}$$

$$\dot{\tau}_c^{-1} \quad = \quad \text{time between scattering events} \tag{2.60}$$

Tight coupling between the photons and baryons occurs when

$$\frac{\dot{\tau}_c^{-1}}{k^{-1}} = \frac{k}{\dot{\tau}_c} \ll 1\ . \tag{2.61}$$

In regime, photons experience so many scattering as they travel across a perturbation that they remain strongly coupled to the baryons.

(a) since the baryon bulk velocity V_b varies on a timescale much longer than $\dot{\tau}_c^{-1}$, show that this implies

$$\Theta_1 \simeq V_b \quad \Leftrightarrow \quad \Theta_2 \simeq 0\ . \tag{2.62}$$

The photon temperature quadrupole, or anisotropic stress, can thus be neglected, which closes the Boltzmann hierarchy. We will hereafter assume the tight-coupling limit, so that we can ignore all multipoles with $\ell \geqslant 2$.

[Acoustic Oscillations]

One generally expands the Boltzmann hierarchy in powers of $k/\dot{\tau}_c$ (the inverse of the optical depth through a wavelength k) and $\omega/\dot{\tau}_c$ (the inverse of the optical depth through a period of oscillation ω). We will remain at first order in $\dot{\tau}_c^{-1}$, which leads to a driven harmonic oscillator equation describing acoustic waves in the photon-baryon fuild. At second order in $\dot{\tau}_c^{-1}$, acoustic oscillations of the monopole and dipole are damped owing to the imperfect coupling between photons and baryons. Photon diffusion creates heat conduction through $\Theta_1 - V_b$ and shear viscosity through Θ_2.

(b) Extract the term $\dot{\tau}_c(\Theta_1 - V_b)$ from Eq.(2.58) and substitute into Eq.(2.56). Show that, after some manipulations, one obtains

$$\big(1+R\big)\ddot{\Theta}_0 + \mathcal{H}R\dot{\Theta}_0 + \frac{k^2}{3}\Theta_0 = -\frac{k^2}{3}\big(1+R\big)\Psi + \mathcal{H}R\dot{\Phi} + \big(1+R\big)\ddot{\Phi}\,. \tag{2.63}$$

Use the fact that $R = 3\bar{\rho}_b/4\bar{\rho}_\gamma \propto a$, i.e. $\dot{R} = \mathcal{H}R$ to reexpress this relation as

$$\frac{d}{d\eta}\Big[\big(1+R\big)\dot{\Theta}_0\Big] + \frac{k^2}{3}\Theta_0 = -\frac{k^2}{3}\big(1+R\big)\Psi + \frac{d}{d\eta}\Big[\big(1+R\big)\dot{\Phi}\Big]\,. \tag{2.64}$$

This is the equation of an oscillator with a time-varying mass $m_{\rm eff} = 1 + R$. The homogeneous equation can be solved by employing the fact that variations over a single period of the oscillation are small.

(c) We have thus far not used Einstein equations. One can show that, in the absence of anisotropic stress (that is, $\pi_\gamma = \pi_\nu = ... = 0$), Einstein equations imply that the two potentials are equal: $\Psi = \Phi$. Neglect the time-dependence of R and show that

$$\frac{d^2}{d\eta^2}\big(\Theta_0 + \Psi\big) + k^2c_s^2\big(\Theta_0 + \Psi\big) = -k^2c_s^2R\Psi + 2\ddot{\Psi}\,, \tag{2.65}$$

where

$$c_s^2 \equiv \frac{\dot{\bar{P}}}{\dot{\bar{\rho}}} \approx \frac{1}{3(1+R)} \tag{2.66}$$

is the adiabatic sound speed of the photon-baryon fluid. c_s defines a characteristic comoving length scale $r_s(\eta)$ known as the *sound horizon*,

$$r_s(\eta) = c_s\int\frac{dt}{a(t)} \equiv c_s\eta\,, \tag{2.67}$$

which represents the comoving distance traveled by a sound perturbation. Calculate r_s when the CMB decouples from the plasma at $\eta_{\rm dec}$ assuming $z_{\rm dec} = 10^3$ and an EdS universe.

(d) Eq.(2.65) is a driven harmonic oscillator equation for the effective temperature $\Theta_0 + \Psi$, in which the gravitational blueshift due to the infall onto a potential well is exactly compensated by Ψ. The frequency $\omega = kc_S$ increasing with decreasing comoving scale k^{-1}. Ignoring the time variation of c_s, R and, especially, $\ddot{\Psi}$, show that the solution to Eq.(2.65) is of the form

$$\Theta_0 + \Psi = -R\Psi + A\cos\big(kc_s\eta\big) + B\sin\big(kc_s\eta\big)\,. \tag{2.68}$$

$\Theta_0 + \Psi$ oscillates around $-R\Psi$ and not zero owing to the baryons, which drag the photons into the potential wells (an effect known as *baryon drag*).

(e) To fix the initial conditions which determine A and B, we take the limit $\eta \to 0$ (or, equivalently, $c_s\eta \equiv r_s \to 0$), in which case (not demonstrated here)

$$\Theta_0 + \Psi \approx \frac{1}{3}\Psi \qquad \text{(adiabatic ICs)} \tag{2.69}$$

$$\Theta_0 + \Psi \approx 2\Psi \qquad \text{(isocurvature ICs)} \tag{2.70}$$

Assuming adiabatic perturbations (generated by inflation for instance) and zero initial "velocity" $\dot{\Theta}_0$, show that the solution is

$$\big(\Theta_0 + \Psi\big)(\eta) = \frac{1}{3}\big(1 + 3R\big)\Psi\cos\big(kc_s\eta\big) - R\Psi\ . \tag{2.71}$$

The baryon drag increases the amplitude of the cosine term. Overall, it accounts for the alternate height of the acoustic peaks (compression peaks occuring at $kc_s\eta = \pi,\ 3\pi,\ ...,$ are enhanced relative to the rarefaction peaks at $kc_s\eta = 2\pi,\ 4\pi,\ ...$) and their enhancement with $R \propto \Omega_b h^2$.

[Integral solution to CMB anisotropies]

Rather than decomposing the Boltzmann equation Eq.(2.37),

$$\frac{d\Theta}{d\eta} = -\hat{n}^i\partial_i\Psi + \dot{\Phi} + \dot{\tau}_c\Big[-\Theta + \frac{1}{4}\delta_\gamma + \hat{\mathbf{n}}\cdot\mathbf{v}_b\Big]\ , \qquad \frac{d}{d\eta} \equiv \partial_\eta + \hat{n}^i\partial_i \tag{2.72}$$

in Legendre polynomials and solve for the multipoles $\Theta_\ell(\eta,\mathbf{k})$, it can be formally integrated to yield the CMB temperature anisotropy $\Theta(\eta_0,\hat{\mathbf{n}})$ as seen by an observer at time η_0:

$$\Theta(\eta_0,\hat{\mathbf{n}}) = \int_0^{\eta_0} d\eta\, e^{-\tau(\eta)}\frac{d\Theta}{d\eta} \qquad \text{where} \qquad \tau(\eta) = \int_\eta^{\eta_0} d\eta'\,\dot{\tau}_c(\eta')\ . \tag{2.73}$$

Here, $\tau(\eta)$ is the (average) optical depth along the line of sight.

(f) To see this, show that

$$\begin{aligned}\Theta(\eta_0,\hat{\mathbf{n}}) &= \int_0^{\eta_0} d\eta\, e^{-\tau}\left(\frac{d\Theta}{d\eta} + \dot{\tau}_c\Theta\right) \\ &= \int_0^{\eta_0} d\eta\, e^{-\tau}\left[-\hat{n}^i\partial_i\Psi + \dot{\Phi} + \dot{\tau}_c\Big(\frac{1}{4}\delta_\gamma + \hat{\mathbf{n}}\cdot\mathbf{v}_b\Big)\right]\ .\end{aligned} \tag{2.74}$$

Next, demonstrate that

$$\int_0^{\eta_0} d\eta\, e^{-\tau}\big(-\hat{n}^i\partial_i\Psi\big) = -e^{-\tau}\Psi\Big|_0^{\eta_0} + \int_0^{\eta_0} d\eta\, e^{-\tau}\Big(\dot{\tau}_c\Psi + \dot{\Psi}\Big) \tag{2.75}$$

Argue that the first term in the right-hand side can be neglected, and substitute this result into the expression of $\Theta(\eta_0,\hat{\mathbf{n}})$ to obtain

$$\Theta(\eta_0,\hat{\mathbf{n}}) = \int_0^{\eta_0} d\eta\, e^{-\tau}\dot{\tau}_c\left(\frac{1}{4}\delta_\gamma + \Psi + \hat{\mathbf{n}}\cdot\mathbf{v}_b\right) + \int_0^{\eta_0} d\eta\, e^{-\tau}\big(\dot{\Psi} + \dot{\Phi}\big)\ . \tag{2.76}$$

Argue that the visibility function

$$g(\eta) \equiv \dot{\tau}_c e^{-\tau} = -\frac{d\tau}{d\eta}e^{-\tau} \tag{2.77}$$

is sharply peaked around the decoupling or last scattering epoch $\eta_{\rm dec}$ to approximate $\Theta(\eta_0,\hat{\mathbf{n}})$ as

$$\Theta(\eta_0,\hat{\mathbf{n}}) \approx \left(\frac{1}{4}\delta_\gamma + \Psi + \hat{\mathbf{n}}\cdot\mathbf{v}_b\right)(\eta_{\rm dec},\hat{\mathbf{n}}) + \int_0^{\eta_0} d\eta\,\big(\dot{\Psi} + \dot{\Phi}\big) \tag{2.78}$$

The first encode the contributions of the intrinsic photon density perturbation ($\frac{1}{4}\delta_\gamma$), Sachs-Wolfe effect from the gravitational potential (Ψ) and Doppler effect from the photon-baryon relative motion ($\hat{\mathbf{n}}\cdot\mathbf{v}_b$). The second term is the integrated Sachs-Wolfe (ISW) effect, which vanishes for time-independent potentials.

Chapter 3

Cosmologies with (massive) neutrinos

3.1 Neutrino timeline

3.1.1 Photon and neutrino density

Below 1 MeV, the relativistic radiation consists of photons and neutrinos, and its energy density is thus generally given by

$$\rho_r = \frac{\pi^2 T_\gamma^4}{15}\left[1+\frac{7}{8}\left(\frac{T_\nu}{T_\gamma}\right)^4 N_\nu\right] . \tag{3.1}$$

Now, if the neutrinos are massless, then both T_γ and T_ν redshift away as $\propto a^{-1}$. Taking the present-day temperature of photons and neutrinos to be $T_{\gamma,0}$ and $T_{\nu,0}$, we have

$$\rho_r = \frac{\pi^2 T_{\gamma,0}^4}{15}\left[1+\frac{7}{8}\left(\frac{T_{\nu,0}}{T_{\gamma,0}}\right)^4 N_\nu\right] a^{-4} . \tag{3.2}$$

However, if neutrinos are massive - neutrino oscillations experiments indicate that $\sum_\nu m_\nu \gtrsim$ 0.06 eV - then the scaling $T_\nu \propto a^{-1}$ is only valid at early time when $T_\nu \gg m_\nu$. We will come back to this shortly.

3.1.2 Neutrino freeze-out

Neutrinos are coupled to the relativistic electron-positron plasma, hence to radiation, via the weak interaction processes

$$\begin{aligned} \nu_e + \bar{\nu}_e &\rightleftharpoons e^+ + e^- \\ e^\pm + \nu_e &\to e^\pm + \nu_e \\ e^\pm + \bar{\nu}_e &\to e^\pm + \bar{\nu}_e . \end{aligned} \tag{3.3}$$

At temperatures much below the scale of electroweak symmetry breaking, $T \ll m_{Z,W^\pm} \sim 100$ GeV, we can determine the cross section of processes mediated by the weak force within the 4-fermion theory of weak interaction. In this limit, the cross section of different weak processes is identical and given by

$$\sigma_F \sim G_F^2 T^2 . \tag{3.4}$$

Since the particles involved are relativistic fermions, their number of d.o.f. is $g_F = 2\times3+2\times2 = 10$ and their number density is $n_F(T) = (3/4)(\zeta(3)/\pi^2)g_F T^3 \simeq 1.3T^3$. Furthermore, setting $v \sim 1$, we find that the interaction rate is

$$\Gamma_F = n_F \langle \sigma_F v \rangle \simeq 1.3 G_F^2 T^5 \; . \tag{3.5}$$

Comparing Γ_F with the expansion rate H, we find

$$\frac{\Gamma_F}{H} \simeq 0.24 T^3 m_P G_F^2 \simeq \left(\frac{T}{1.4\,\mathrm{MeV}}\right)^3 \; . \tag{3.6}$$

At temperature $T \lesssim 1.4$ MeV, the probability for a neutrino to interact within one hubble time H^{-1} is much less than unity and the neutrinos are effectively decoupled. The heat bath becomes transparent to neutrinos which are no longer in thermal equilibrium with electrons and positrons, hence photons and baryons.

Two comments:

- The number densities of μ- and τ-leptons are negligibly small at MeV-temperatures. Therefore, the only reactions enforcing thermal contact between μ- and τ-neutrinos and the plasma are the elastic scattering

$$\begin{aligned} e^\pm + \nu_{\mu,\tau} &\to e^\pm + \nu_{\mu,\tau} \\ e^\pm + \bar{\nu}_{\mu,\tau} &\to e^\pm + \bar{\nu}_{\mu,\tau} \; . \end{aligned} \tag{3.7}$$

  These are entirely due to Z-boson exchange. Consequently, the cross-section $\propto m_Z^{-2}$ (with $m_Z \approx 91.19$ GeV) is smaller than the total-cross section of $e^\pm$-ν_e interactions, and the μ- and τ-neutrino decouple earlier than the electron neutrinos.

- The chemical potential of electrons and positrons is usually neglected since one expects $\mu_e/T \ll 1$. The interaction

$$e^+ + e^- \rightleftharpoons \nu_e + \bar{\nu}_e \tag{3.8}$$

  thus implies $\mu_{\nu_e} \simeq -\mu_{\bar{\nu}_e}$. Unfortunately, the number $n_{\nu_e} - n_{\bar{\nu}_e}$ which determines, together with $n_{e^-} - n_{e^+}$, the lepton number of the Universe is not known from observations. Therefore, one ussually assumes that, like the baryon number, the lepton number is small so that one can also neglect the chemical potential of neutrinos.

3.1.3 Neutrino temperature

The neutrino temperature today is actually less than that of the photons because neutrinos *decoupled* at $T \simeq 1.4$ MeV before electrons e^- and positrons e^+ annihilated at $T \simeq 0.5$ MeV. The entropy of electrons and positrons was transferred to the photons during $e^\pm$ annihilation, since the latter produced photons and (nearly) no neutrinos. To derive the final photon and neutrino temperature, let us work out the effective number of d.o.f. in entropy, g_{*s}, for the photon - $e^\pm$ component before and after $e^\pm$ annihilation.

Since the entropy density s is dominated by the contribution from relativistic particles, we have in a very good approximation

$$s = \frac{2\pi^2}{45} g_{*s} T^3 \; , \tag{3.9}$$

where

$$g_{*s}(T) = \sum_{\text{bosons}} g_X \left(\frac{T_X}{T}\right)^3 + \frac{7}{8} \sum_{\text{fermions}} g_X \left(\frac{T_X}{T}\right)^3 . \tag{3.10}$$

Here, T_X is the temperature of species X while T is the temperature of the photon bath by definition. For the problem at hand,

$$g_{*s}(T_i) = 2 + \frac{7}{8} \times 4 = \frac{11}{2} \tag{3.11}$$

$$g_{*s}(T_f) = 2 , \tag{3.12}$$

where $T_i > m_e$ and $T_f < m_e$ are the initial and final photon temperatures. Since the entropy of photons and $e^\pm$ in equilibrium is separately conserved (as is the entropy of decoupled species like neutrinos), we must have

$$g_{*s}(T_i) a_i^3 T_i^3 = g_{*s} a_f^3 T_f^3 \Leftrightarrow a_f^3 T_f^3 = \frac{11}{4} a_i^3 T_i^3 \tag{3.13}$$

Therefore, since $T_i \equiv T_\nu$ and $T_f \equiv T_\gamma$ and since the temperatures all scale like a^{-1}, it follows that

$$T_\gamma / T_\nu = (11/4)^{1/3} \tag{3.14}$$

after $e^\pm$ annihilation. The entropy of neutrino sea and photon sea are separately conserved.

With this result, we can determine the ratio of neutrino to photon density,

$$\frac{\rho_\nu}{\rho_\gamma} = \frac{7}{8} \left(\frac{4}{11}\right)^{4/3} N_\nu = 0.2271 N_\nu . \tag{3.15}$$

The number of neutrinos is sometimes left as a parameter even though accelerators have established that $N_\nu = 3$. One reason is that non-standard cosmologies could modify the value of ρ_ν/ρ_γ given here, e.g. the presence of an additional massless particle that decoupled at very high energies could mimic a neutrino. Another reason is that $e^\pm$ annihilation begins before neutrino decoupling has completed, and because interactions modify the free-particle thermodynamics assumed here. For all these reasons, one usually allows N_ν to vary. So far, $N_\nu \approx 3.04$ (Planck CMB experiment).

3.1.4 Phase space conservation

Consider the distribution function f_X of a particle which has *frozen out* or *decoupled*, i.e. which has stopped interacting, at a conformal time η_D corresponding to a temperature T_D. Phase space conservation implies that f_X obeys at all time $\eta \geqslant \eta_D$ the collisionless Liouville equation

$$\partial_\eta f_X - \mathcal{H} p \partial_p f_X = 0 , \tag{3.16}$$

where $p \equiv \sqrt{g_{ij} p^i p^j}$ (with $g_{ij} = a^2 \gamma_{ij}$) is again the modulus of the physical 3-momentum. This implies that $f_X \equiv f_X(ap)$ only depends on ap, which is nothing but another formulation of the redshift of momenta. After the *freeze-out* or *decoupling*, we can thus write $f_X(p)$ as

$$f_X(p) = f_{\text{eq}}\left(p \frac{a(\eta)}{a(\eta_D)}, T_D\right) . \tag{3.17}$$

Here, f_{eq} is the distribution function describing the thermal equilibrium immediately before the decoupling. Note that, after decoupling, the temperature is not strictly speaking a genuine thermodynamic temperature (since there are no interactions), but merely a parameter that describes the phase space distribution function.

3.1.5 Background of massive relic neutrinos

For relativistic fermions at decoupling like neutrinos (i.e. $T_D \gg m_X$), equation (3.17) yields ($a_D \equiv a(\eta_D)$)

$$f_X(p) = \frac{1}{\exp\left(\frac{ap}{a_D T_D}\right) + 1} \tag{3.18}$$

after freeze-out, which implies that $T_X \propto a^{-1}$ for $\eta \geqslant \eta_D$.

After decoupling, their number density satisfies $n_\nu \propto a^{-3}$ and their phase space density f_ν remains Fermi-Dirac at all time $\eta \geqslant \eta_D$, i.e. even after they became non-relativistic at a temperature $T \leqslant m_\nu \ll 1$ MeV. Although the neutrinos are not described by a thermal distribution $\rho_\nu \propto T_\nu^4$ for $T_\nu \leqslant m_\nu$, we can still define $T_{\nu,0} \equiv T_\nu a \equiv T_{\nu,D} a_D$ *at all time* $\eta \geqslant \eta_D$ and consider it to be a *temperature parameter* rather than a physical temperature.

Once the effective temperature $T_{\nu,0}/a$ drops below $\lesssim m_\nu$, the neutrino density and pressure must be calculated from:

$$\rho_\nu = \sum_{\nu=1}^{N_\nu} \frac{2}{(2\pi)^3} \int_0^\infty dp\, 4\pi p^2 \sqrt{p^2 + m_\nu^2} f_\nu(p,a) \tag{3.19}$$

$$P_\nu = \sum_{\nu=1}^{N_\nu} \frac{2}{3(2\pi)^3} \int_0^\infty dp\, 4\pi p^2 \frac{p^2}{\sqrt{p^2 + m_\nu^2}} f_\nu(p,a) \ , \tag{3.20}$$

where $f_\nu(p,a) = (e^{ap/T_{\nu,0}} + 1)^{-1}$ (since we ignore the possibility of a non-zero neutrino chemical potential) and p denotes the physical momentum.

The present-day energy density of this unobservable relic neutrino background is

$$\Omega_\nu = \frac{\sum m_\nu}{93.6 h^2 \text{ eV}} \ . \tag{3.21}$$

while its temperature is $T_{\nu,0} = (4/11)^{1/3} T_{\gamma,0} \approx 1.96$K for a CMB temperature $T_{\gamma,0} = 2.725$K. It is made up of neutrinos and anti-neutrinos in unknown proportion since the lepton number of the Universe is not known.

- For photons, decoupling occurs at recombination, $z \sim 10^3$, after which the blackbody shape of photon distribution remains unchanged but with a temperature $T_\gamma \propto 1/a$. Furthermore, photons remain relativistic at all time.

- The redshift of matter-radiation equality is sensitive to $T_{\gamma,0}$ and the neutrino fraction. For the values quoted above, we have

$$1 + z_{\text{eq}} = 2.39 \times 10^4\, \Omega_m h^2 \ . \tag{3.22}$$

  The (primary) CMB anisotropies form shortly after matter-radiation equality.

3.2 Neutrino perturbations

Once the neutrino decouple from the plasma at $T \sim 1.4$ MeV, they become collisionless so that one must evolve their full phase space distribution - like photons - with the important difference that there is no collision term. Therefore, one solves the collisionless Boltzmann or Vlasov equation.

Therefore, neutrinos - like photons - require more information to characterize: they have monopole, dipole but also higher multipoles (quadrupole etc.) perturbations. In analogy with the treatment for photon, one introduces a fractional perturbation $\mathcal{N}(\eta, \mathbf{k}, \hat{\mathbf{n}})$ to the neutrino distribution function through the relation

$$f_\nu \approx \bar{f}_\nu - p\frac{\partial \bar{f}_\nu}{\partial p}\mathcal{N} \ , \tag{3.23}$$

and perform a Legendre decomposition starting from the Boltzmann equation for massive particles. The multipole moments of the neutrino are denoted as $\mathcal{N}_\ell$.

The neutrino quadrupole $\mathcal{N}_2$ reflects the fact that neutrino can free-stream from regions with a larger temperature to regions with a lower temperature. At the level of an (effective) fluid description, this effect manifests itself as an anisotropic stress $\subset \pi_{\mu\nu}$ proportional to the neutrino "*shear viscosity*" π_ν. Let us see how $\mathcal{N}_2$ and π_ν are related.

3.2.1 Perturbed energy-momentum tensor

To calculate the energy-momentum tensor of neutrinos, we must take into account the fact that neutrinos are generally not at rest in our coordinate frame. In other words, there is a neutrino energy current $j^\mu = \rho_\nu u^\mu$ with non-vanishing spatial components. The 4-velocity u^μ of energy transport is, at first order in perturbations,

$$u^\mu = \frac{1}{a}\big(1 - \Psi, v^i\big) \ , \qquad v^i = \frac{dx^i}{d\eta} \ . \tag{3.24}$$

At linear PT order, v^i can be identified with the proper peculiar velocity. This implies

$$u_\nu = \big(-a(1+\Psi), v_i\big) \ . \tag{3.25}$$

Consider now the mixed indices T^μ_ν to reduce the presence of metric fluctuations. Substituting this expression into Eq.(1.23) and using the longitudinal gauge, we find:

$$\begin{aligned} T^0_0 &= -\rho \\ T^i_0 &= -\big(\rho + P\big)v^i \\ T^0_i &= \big(\rho + P\big)v_i \\ T^i_j &= P\delta^i_j + \pi^i_j \ . \end{aligned} \tag{3.26}$$

The energy-momentum tensor can also be decomposed in scalar, vector and tensor d.o.f. There are four scalar d.o.f.: ρ, P, $v^{\parallel}_i$, $\pi^{\parallel}_{ij}$. The last one is a scalar-type anisotropic stress, which can be constructed out of a potential $\pi^{\parallel}$:

$$\pi^{\parallel}_{ij}(\eta, \mathbf{x}) = \Big(\partial_i\partial_j - \frac{1}{3}\delta_{ij}\nabla^2\Big)\pi^{\parallel}(\eta, \mathbf{x}) \ . \tag{3.27}$$

In Fourier space, this becomes

$$\pi^{\parallel}_{ij}(\eta, \mathbf{k}) = -k^2\Big(\hat{\mathbf{k}}_i\hat{\mathbf{k}}_j - \frac{1}{3}\delta_{ij}\Big)\pi^{\parallel}(\eta, \mathbf{k}) \ . \tag{3.28}$$

It is convenient to define $\pi^{\|}$ as

$$\pi^{\|}(\eta, \mathbf{k}) \equiv k^{-2} \bar{P}_\nu \pi_\nu(\eta, \mathbf{k}) \, , \tag{3.29}$$

so that the neutrino shear viscosity $\pi_\nu(\eta, \mathbf{k})$ is dimensionless and can be directly compared to the quadrupole $\mathcal{N}_2(\eta, \mathbf{k})$.

3.2.2 Neutrino anisotropic stress

Let $f(\eta, x^i, p, \hat{n}^i)$ be the phase space distribution of a particle subject to the on-shell condition $E^2 = p^2 + m^2$, where E and $p = \sqrt{g_{ij} p^i p^j}$ are the physical energy and momentum (i.e. they scale as $\propto a^{-1}$ for freely propagating particles).

The covariant expression of the energy-momentum tensor is

$$T^\mu_\nu = \frac{g_s}{(2\pi)^3} \int \prod_{i=1}^{3} dp_i \, (-g)^{-1/2} \, \frac{p^\mu p_\nu}{p^0} \, f(\eta, x^i, p, \hat{n}^i) \, , \tag{3.30}$$

where g_s is the spin degeneracy of the particle, and g is the determinant of the metric $g_{\mu\nu}$.

In the conformal Newtonian gauge, we have $(-g)^{-1/2} = a^{-4}\big(1 - \Psi + 3\Phi\big)$. Furthermore, the components of the coordinate momentum of a particle are

$$p^0 = \frac{E}{a}\big(1 - \Psi\big) \, , \qquad p^i = \frac{p}{a}\big(1 + \Phi\big)\hat{n}^i \tag{3.31}$$

$$p_0 = -aE\big(1 + \Psi\big) \, , \qquad p_i = ap\big(1 - \Phi\big)\hat{n}_i \tag{3.32}$$

at first order in perturbations, and the volume measure reads $dp_1 dp_2 dp_3 = a^3\big(1 - 3\Phi\big) d^3p = a^3\big(1 - 3\Phi\big) p^2 dp d\Omega$, where Ω is the solid angle characterizing the direction of $\hat{\mathbf{n}}$.

Let us now specialize these expressions to relativistic neutrinos. $g_s = 2$ for each neutrino family. The unperturbed components of the neutrino energy-momentum tensor $\overline{T}^\mu_\nu(\eta)$ are given by

$$\overline{T}^{00} = -\frac{g_s}{(2\pi)^3} \int d\Omega \, dp \, p^3 \, \bar{f}_\nu = -\bar{\rho}_\nu \tag{3.33}$$

$$\overline{T}^0_i = \overline{T}^i_0 = 0 \tag{3.34}$$

$$\overline{T}^{ij} = \frac{g_s}{(2\pi)^3} \int d\Omega \, dp \, p^3 \, \hat{n}^i \hat{n}_j \, \bar{f}_\nu = \frac{1}{3} \bar{\rho}_\nu \delta^i_j \equiv \bar{P}_\nu \delta^{ij} \, . \tag{3.35}$$

Note that $\overline{T}^i_j = \bar{P}_\nu \delta^i_j$ indicates that the stress is isotropic for the homogeneous background. Hence, $\overline{T}^\mu_\nu$ is diagonal as expected.

Consider now the perturbation $\delta T^{\mu\nu}(\eta, \mathbf{x})$ to the energy-momentum tensor. At first order in perturbations, the perturbation to the neutrino energy density is given by

$$\begin{aligned} \delta T^0_0 &= -\frac{g_s}{(2\pi)^3}\int d\Omega\, dp\, p^3\, \delta f_\nu \\ &= -\frac{g_s}{(2\pi)^3}\int d\Omega\, dp\, p^3 \left(-p\frac{\partial \bar{f}_\nu}{\partial p}\mathcal{N}\right) \\ &= \frac{g_s}{(2\pi)^3}\int d\Omega \left\{ p^4 \mathcal{N} \bar{f}_\nu \Big|_0^\infty - \int dp\, \bar{f}_\nu \frac{\partial}{\partial p}\left(p^4\mathcal{N}\right) \right\} \\ &= -4\bar{\rho}_\nu \frac{1}{4\pi}\int d\Omega\, \mathcal{N} \\ &= -4\bar{\rho}_\nu \mathcal{N}_0 \;, \end{aligned} \tag{3.36}$$

Analogously, the perturbation to the neutrino energy flux along the i direction is

$$\delta T^0_i = 4\bar{\rho}_\nu \frac{1}{4\pi}\int d\Omega\, \hat{n}^i\, \mathcal{N} \;. \tag{3.37}$$

To evaluate the integral over the unit vector $\hat{\mathbf{n}}$, we Fourier transform this relation and consider the contraction $\hat{k}^i \delta T^0_i(\eta, \mathbf{k})$:

$$\begin{aligned} \hat{k}^i \delta T^0_i(\eta,\mathbf{k}) &= 4\bar{\rho}_\nu \frac{1}{4\pi}\int d\Omega\, (\hat{\mathbf{k}}\cdot\hat{\mathbf{n}})\, \mathcal{N}(\eta,\mathbf{k},\hat{\mathbf{n}}) \\ &= 4\bar{\rho}_\nu \frac{1}{2}\int d\mu\, \mu\, \mathcal{N}(\eta,\mathbf{k},\hat{\mathbf{n}}) \\ &= -i\frac{4}{3}\bar{\rho}_\nu \mathcal{N}_1(\eta,\mathbf{k}) \;. \end{aligned} \tag{3.38}$$

This shows that

$$\delta T^0_i(\eta,\mathbf{k}) = -i\hat{k}_i \frac{4}{3}\bar{\rho}_\nu \mathcal{N}_1(\eta,\mathbf{k}) = -i\hat{k}_i\big(\bar{\rho}_\nu + \bar{P}_\nu\big)\mathcal{N}_1(\eta,\mathbf{k}) \;. \tag{3.39}$$

which can be Fourier transformed back to obtain $\delta T^0_i(\eta, \mathbf{x})$. This shows that $-i\hat{k}_i \mathcal{N}_1(\eta, \mathbf{k})$ is the Fourier transform of the energy flux in the i-direction.

Finally, we have

$$\begin{aligned} \delta T^i_j &= \frac{g_s}{(2\pi)^3}\int d\Omega\, dp\, p^3\, \hat{n}^i \hat{n}_j\, \delta f_\nu \\ &= 4\bar{\rho}_\nu \frac{1}{4\pi}\int d\Omega\, \hat{n}^i \hat{n}_j\, \mathcal{N} \;. \end{aligned} \tag{3.40}$$

We take the Fourier transform of this relation to obtain $\delta T^i_j(\eta, \mathbf{k})$, and consider the following two contractions:

$$\begin{aligned} \hat{k}_i \hat{k}^j \delta T^i_j(\eta,\mathbf{k}) &= 4\bar{\rho}_\nu \frac{1}{4\pi}\int d\Omega\, (\hat{\mathbf{k}}\cdot\hat{\mathbf{n}})^2\, \mathcal{N}(\eta,\mathbf{k},\hat{\mathbf{n}}) \\ &= 4\bar{\rho}_\nu \frac{1}{2}\int d\mu\, \mu^2\, \mathcal{N}(\eta,\mathbf{k},\hat{\mathbf{n}}) \\ &= \frac{4}{3}\bar{\rho}_\nu \left[\mathcal{N}_0(\eta,\mathbf{k}) - \frac{2}{5}\mathcal{N}_2(\eta,\mathbf{k})\right] \;. \end{aligned} \tag{3.41}$$

and

$$\delta^j_i \delta T^i_j(\eta, \mathbf{k}) = 4\bar{\rho}_\nu \mathcal{N}_0(\eta, \mathbf{k}) \ . \tag{3.42}$$

Owing to their index dependence, the perturbed components δT^i_j are necessarily of the form

$$\begin{aligned}\delta T^i_j(\eta, \mathbf{k}) &= t_1(\eta, \mathbf{k})\delta^i_j + t_2(\eta, \mathbf{k})\left(\hat{k}^i\hat{k}^j - \frac{1}{3}\delta^{ij}\right) \\ &\equiv \delta P_\nu(\eta, \mathbf{k})\delta^i_j - \bar{P}_\nu \pi_\nu(\eta, \mathbf{k})\left(\hat{k}^i\hat{k}_j - \frac{1}{3}\delta^i_j\right) \ .\end{aligned} \tag{3.43}$$

The first term contributes to the isotropic stress, whereas the second contributes to the anisotropic stress. Therefore, δP_ν is the perturbation to the neutrino pressure and π_ν is the effective neutrino shear viscosity. Using Eqs.(3.41) and (3.42), we find $\delta P_\nu = 4\bar{P}_\nu \mathcal{N}_0$, as expected, and

$$\pi_\nu = \frac{12}{5}\mathcal{N}_2 \ . \tag{3.44}$$

This shows that the neutrino quadrupole is directly proportional to the neutrino anisotropic stress.

From this point onwards, one proceeds as in the case of photons. However, it is not possible to truncate the hierarchy for neutrinos (at least so long as they are relativistic) because of their large free-streaming.

3.2.3 Neutrino free-streaming

Even though sound waves cannot propagate in the collisionless neutrino component, it is possible to define a characteristic distance over which neutrinos free-stream. This (physical) *free-streaming length* $\lambda_{\rm FS}$ is defined analogously to the Jeans scale:

$$\lambda_{\rm FS} = 2\pi \frac{a(t)}{k_{\rm FS}} \ , \qquad k^2_{\rm FS}(t) = \frac{4\pi G\bar{\rho}(t)a^2}{\sigma^2_v} = \frac{3}{2}\frac{a^2H^2}{\sigma^2_v} \ , \tag{3.45}$$

where σ_v is a characteristic velocity dispersion. Here, $k_{\rm FS}$ is the *comoving* free-streaming wavenumber.

So long as the neutrinos are relativistic, they travel approximately at the speed of light, so that their free-streaming basically is the Hubble radius, $\lambda_{\rm FS} \propto H^{-1} \propto t$. Once they become non-relativistic at $a = a_{\rm nr}$, their thermal velocity dispersion scales like

$$\sigma^2_v(t) = \frac{\langle p^2\rangle}{m^2_\nu} \simeq 3.6\times 10^{-7}\left(\frac{1\ {\rm eV}}{m_\nu}\right)^2 \frac{1}{a^2} \ . \tag{3.46}$$

For $a > a_{\rm nr}$, their free-streaming scale thus grows like

$$\lambda_{\rm FS} \propto t^{1/3} \tag{3.47}$$

in matter domination, more slowly than the scale factor $a \propto t^{2/3}$. As a consequence, the comoving free-streaming wavenumber $k_{\rm FS}$ reaches a minimum at $a = a_{\rm nr}$:

$$k_{\rm nr} \equiv k_{\rm FS}(a_{\rm nr}) \approx 0.018\,\Omega^{1/2}_m \left(\frac{m_\nu}{1\ {\rm eV}}\right)^{1/2} \ h^{-1}{\rm Mpc} \ . \tag{3.48}$$

This implies

- For $k < k_{\rm nr}$, the neutrino velocity prior to the non-relativistic transition can be considered as vanishing: the neutrino distribution is static, with a local density (and temperature) perturbation proportional to the metric fluctuations. After the non-relativistic transition, the neutrino perturbations behave like cold dark matter perturbations.

- For $k > k_{\rm nr}$, free-streaming damps small-scale neutrino density fluctuations: neutrinos cannot be confined into regions smaller than the free-streaming length, for obvious kinematic reasons. The metric perturbations are also damped on those scales by gravitational backreaction. The multipoles are populated one after each other, at the expense of the low multipoles which get reduced. Physically, this corresponds to the fact that an observer locally sees the superposition of a growing number of neutrino flows, coming from all directions. Since the flows come from a random distribution of over- and under-densities, the local density contrast tends to average to zero in the limit $k \gg k_{\rm nr}$.

- Once the neutrino becomes non-relativistic, the (comoving) free-streaming scale drops so that a fluid description is a good approximation at wavenumbers $k < k_{\rm FS}(t)$.

3.2.4 Growth of linear perturbations

After the non-relativistic transition, the neutrinos can be approximated by a fluid with the free-streaming scale $k_{\rm FS}$ playing the role of the Jeans scale. Introducing the relative contribution

$$f_\nu = \frac{\Omega_\nu}{\Omega_m} \tag{3.49}$$

of massive neutrinos to the total matter density, the continuity and Euler equations for the CDM and non-relativistic neutrinos can be combined into

$$\begin{aligned} \ddot{\delta}_c + \mathcal{H}\dot{\delta}_c &= \frac{3}{2}\mathcal{H}^2\Big[f_c\delta_c + f_\nu\delta_\nu\Big] \\ \ddot{\delta}_\nu + \mathcal{H}\delta_\nu &= \frac{3}{2}\mathcal{H}^2\Big[f_c\delta_c + \big(f_\nu - k^2/k_{\rm FS}^2\big)\delta_\nu\Big] \end{aligned} \tag{3.50}$$

where $f_c = 1 - f_\nu$ is the fraction of mass density in CDM. The Poisson equation was used to obtain the right-hand side. Changing the time coordinates from $\eta \mapsto y \equiv \log(a/a_{\rm nr})$, these coupled differential equations reduce to

$$\begin{aligned} \delta_c'' + \frac{1}{2}\delta_c' &= \frac{3}{2}\Big[f_c\delta_c + f_\nu\delta_\nu\Big] \\ \delta_\nu'' + \frac{1}{2}\delta_\nu' &= \frac{3}{2}\Big[f_c\delta_c + f_\nu\delta_\nu - e^{\tau}\kappa^2\delta_\nu\Big] \;. \end{aligned} \tag{3.51}$$

in matter domination. Here, $\kappa = k/k_{\rm nr}$ and a prime denotes a derivative w.r.t. y. Both δ_c and δ_ν are now functions of (y, k).

A simple solution to the time-dependent growth of $\delta_c(y, k)$ can be obtained as follows: on scales much less than the neutrino free-streaming scale, the evolution of δ_c obeys the growth equation obtained by setting the neutrino density for the reason mentioned above . Therefore, the CDM growing mode satisfies

$$\delta_c^{(+)}(y, k) = e^{r_+ y} \tag{3.52}$$

on scales $k \gg k_{\mathrm{FS}}$, with

$$r_+ = \frac{1}{4}\left(-1 + \sqrt{25 - 24 f_\nu}\right) \approx 1 - \frac{3}{5} f_\nu \ . \tag{3.53}$$

The last equality assumes $f_\nu \ll 1$. The growth of small-scale CDM perturbations is, therefore, suppressed relative to large scales, and we expect this to be reflected in the distribution of bound structures such as dark matter halos.

Chapter 4

The growth of large scale structure

Analytic techniques for understanding structure formation are accurate only in the weakly non-linear regime $\delta \lesssim 1$. To probe the strongly nonlinear regime $\delta \gg 1$ such as the deep core of virialized structures, one must resort to numerical simulations. Nevertheless, it is possible to get analytic solutions that are valid deep into the nonlinear regime provided one makes simplifying assumptions.

4.1 The spherical collapse model

One such toy model is the spherical collapse model. For simplicity, we specialize to a flat background Universe of density $\bar{\rho}(t) = \rho_{cr}(t)$. For a matter-dominated Universe, the critical solution is the Einstein de-Sitter (EdS) Universe with scale factor $a \propto t^{2/3}$ and density $1/(6\pi G t^2)$. We assume that, *at some early time* $t_i \ll 1$, there is a tophat density perturbation of proper radius R_i centered at the origin of our coordinate system. The density enclosed within R_i is $[\rho_{cr}(t_i)(1+\delta_i(R_i)]$, whereas the mass is

$$M = \frac{4\pi R_i^3}{3}\rho_c(t_i)\,(1+\delta_i)\ . \tag{4.1}$$

We assume $0 < \delta(t_i) \ll 1$ such that we recover the linear growth at first. Due to the enhanced gravitational force, the perturbation will expand slightly more slowly than the Universe as a whole before coming to a halt and recollapsing.

Note that, to remain a bit more general, one could also have considered a spherically symmetric density perturbation with density contrast monotonically increasing with decreasing radius. In General Relativity, Birkhoff's theorem indeed states that any spherically symmetric region of space-time will evolve independently of the surrounding space-time. This implies that each concentric shell remains concentric and evolves like a closed ($\delta > 0$) or open ($\delta < 0$) Friedman Universe, independently of the background and the outer mass distribution.

4.1.1 Parametric solution

For a closed Universe, the Friedmann equations yield a parametric solution for the scale factor a and the time t in terms of a development angle ϑ. This solution can be adpated to describe the time evolution of an overdense shell:

$$\frac{R(\vartheta)}{R_i} = A\,(1-\cos\vartheta) \tag{4.2}$$

$$\frac{t(\vartheta)}{t_i} = B\,(\vartheta-\sin\vartheta)\ , \tag{4.3}$$

where R is the physical radius of the perturbation. Note that the evolution of initially underdense regions ($\delta_i < 0$), which we will not study here, is obtained upon replacing $(\vartheta-\sin\vartheta) \to (\sinh\vartheta-\vartheta)$ and $(1-\cos\vartheta) \to (\cosh\vartheta - 1)$.

- Ensuring that we recover the background solution at early time (i.e. $\vartheta \ll 1$) introduces a relation between A and B. In this regime, energy conservation – which follows form the equation of motion – implies that

$$E = K + W = \frac{1}{2}\left(\frac{dR}{dt}\right)^2 - \frac{GM}{R} \approx 0\ , \tag{4.4}$$

  where $E = K + W$ is the total energy of a test particle (of unit mass) moving with the shell. Since the initial velocities are given by the hubble flow, $(dR/dt) \approx HR$ with $H(t) = 2/(3t)$, so that

$$\frac{1}{2}\,(HR)^2 = \frac{2}{9}\left(\frac{R}{t}\right)^2 \equiv \frac{GM}{R}\ . \tag{4.5}$$

- Expanding R and t at leading order in ϑ, we find $R \approx R_i A\vartheta^2/2$ and $t \approx t_i B\vartheta^3/6$. On substituting these expression into the above relation, we eventually obtain

$$(R_i A)^3 = GM\,(Bt_i)^2\ . \tag{4.6}$$

  This is linear growth in EdS (from the background point of view), or Kepler's third law for a radial orbit (from the perturbation point of view).

- Any concentric shell initially expands with the hubble flow and, owing to its self-gravity, reaches a maximum size at $\vartheta = \pi$ before eventually collapsing onto a point at $\vartheta = 2\pi$. The constant A is set by requiring $R_{\rm max}/R_i = 2A$. Let $K_i \equiv (H_i R_i)^2/2$ and $W_i = -GM/R_i$ be the kinetic and potential energy at time $t = t_i$. Conservation of energy implies that

$$E = -\frac{GM}{R_{\rm max}} = K_i + W_i\ , \tag{4.7}$$

  or, equivalently,

$$E = -\frac{GM}{R_{\rm max}} = -\frac{R_i}{R_{\rm max}} K_i\,(1+\delta_i) = K_i - \frac{4\pi G}{3}\rho_{cr}(t_i)\,(1+\delta_i)\,R_i^2 = -K_i\delta_i\ , \tag{4.8}$$

  which shows that collapse occurs only if $\delta_i > 0$. Therefore,

$$\frac{R_{\rm max}}{R_i} = 1 + \frac{1}{\delta_i} = 2A\ . \tag{4.9}$$

  B is then fully determined by Eq.(4.6).

- Defining $\Omega_i = 1+\delta_i \approx 1$ as the initial density within the shell in unit of the critical density, we have

$$A = \frac{\Omega_i}{2\left(\Omega_i - 1\right)} \approx \frac{1}{2\delta_i} \tag{4.10}$$

$$B = \frac{\Omega_i}{2H_i t_i \left(\Omega_i - 1\right)^{3/2}} = \frac{3\Omega_i}{4\left(\Omega_i - 1\right)^{3/2}} \approx \frac{3}{4\delta_i^{3/2}} \; . \tag{4.11}$$

This shows that $B^2/A^3 = 9/2$ in the limit $\delta_i \to 0$, which is achieved by taking $t_i \to 0$.

Eqs. (4.2)–(4.3) take precisely the parametric form of the solution to the Friedmann equation for a matter-dominated Universe with density parameter Ω_i.

4.1.2 Linear regime and turnaround

The ratio of the average density within the shell to that of the background evolves according to

$$1+\Delta = \frac{[M/V(t)]}{\rho_{cr}(t)} \approx \frac{\rho_{cr}(t_i)}{\rho_{cr}(t)} \left(\frac{R_i}{R}\right)^3 = \frac{B^2\left(\vartheta - \sin\vartheta\right)^2}{A^3\left(1-\cos\vartheta\right)^3} = \frac{9}{2}\frac{\left(\vartheta - \sin\vartheta\right)^2}{\left(1-\cos\vartheta\right)^3} \; . \tag{4.12}$$

To derive the behaviour at early time $t \approx t_i$, we expand $\sin\vartheta$ and $\cos\vartheta$ up to order ϑ^5 and find

$$\Delta \approx \frac{3}{20}\vartheta^2 \approx \frac{3}{20}\left(\frac{6t}{Bt_i}\right)^{2/3} \approx \frac{3}{5}\delta_i \left(\frac{t}{t_i}\right)^{2/3} \; , \tag{4.13}$$

where we have used $B \approx 3/(4\delta_i^{3/2})$ to obtain the last equality.

- Note that the factor of $3/5$ arises from the fact that the initial conditions are a mixture of growing and decaying mode. Namely

$$\begin{pmatrix} \delta_i \\ \theta_i \end{pmatrix} \equiv \begin{pmatrix} \delta_i \\ 0 \end{pmatrix} = \frac{3}{5}\delta_i \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{2}{5}\delta_i \left(\; 1// - \tfrac{2}{3} \; \right) \; . \tag{4.14}$$

Here, $\theta = -\boldsymbol{\nabla}\cdot\mathbf{u}/\dot{a}$ is the divergence of the linear peculiar velocity in units of $-1/\dot{a}$. Thus, the spherical collapse reduces to the linear theory growth law $D(a) \propto a$ of a growing mode of amplitude $(3/5)\delta_i$ when the initial peculiar velocity vanishes.

Using Eq.(4.13), the linear theory prediction at any time t reads

$$\delta_{\rm L} = \frac{3}{5}\delta_i \left(\frac{t}{t_i}\right)^{2/3} = \frac{3}{5}\left(\frac{3}{4}\right)^{2/3}\left(\vartheta - \sin\vartheta\right)^{2/3} \; . \tag{4.15}$$

At $\vartheta = \pi$, when the shell stops expanding and separate out of the expanding background, the linear overdensity is $\delta_{\rm L} = (3/5)(3\pi/4)^{2/3} = 1.062$. This should be compared to the actual density within the shell relative to the background

$$1+\Delta_{\rm max} = \frac{9\pi^2}{16} \approx 5.55 \; . \tag{4.16}$$

Turnaround thus corresponds to the time at which linear theory become invalid: the object is already significantly nonlinear. The departure from linear theory further increases as the collapse proceeds.

4.1.3 Virialization

At $\vartheta = 2\pi$, the nonlinear solution formally collapses to a point, i.e. $\Delta \to \infty$. In practice however, a perturbation is not spherical and – owing to internal pressure gradients (gas), angular momentum conservation etc. (dust) – will virialize and form a finite-size bound object. The final dimensions of the virialized system can be worked out by the following argument.

Virial equilibrium states that $2K_{\rm vir}+W_{\rm vir} = 0$. Furthermore, the total energy $E = K_{\rm vir}+W_{\rm vir}$ must be the same as that at turnaround. Therefore,

$$K_{\rm vir} + W_{\rm vir} = \frac{W_{\rm vir}}{2} \approx -\frac{GM}{2R_{\rm vir}} = -\frac{GM}{R_{\rm max}} \; . \tag{4.17}$$

This shows that $R_{\rm vir} \approx R_{\rm max}/2$, i.e. the density within the shell increased by an additional factor of 8 while the background density decreased. Assuming that $t_{\rm vir} = t(\vartheta = 2\pi)$ implies that $t_{\rm vir} = 2t_{\rm max}$. In our EdS Universe, the scale factor thus increased by $2^{2/3}$ between $t_{\rm max}$ and $t_{\rm vir}$, while the background density went down by a factor of 4. Therefore, the density at virialization is

$$1 + \Delta_{\rm vir} = \frac{9\pi^2}{16}\left(\frac{R_{\rm max}}{R_{\rm vir}}\right)^3 \left(\frac{\rho_{cr}(t_{\rm max})}{\rho_{cr}(t_{\rm vir})}\right) = \frac{9\pi^2}{16} \times 8 \times 4 \approx 178 \tag{4.18}$$

times densier than the background. This also suggests that all virialized objects will have the same density relative to the background, regardless of their mass. This relation forms the basis for the statement that the virialized region of, e.g., a cluster is a sphere with average density ~ 200 times the critical density of the Universe at the collapse epoch.

- This threshold is often used to defined virialized mass concentrations of dark matter, the so-called *dark matter halos*, in numerical simulations. Another consequence of this relation is that a massive cluster with a comoving virial radius of $1.5\ h^{-1}{\rm Mpc}$ must have formed form the collapse of a region with comoving radius $1.5 \times 178^{1/3} \sim 10\ h^{-1}{\rm Mpc}$.

- We can obtain the overdensity predicted by linear theory at the time of virialization upon setting $\vartheta = 2\pi$ in Eq.(4.15),

$$\delta_{\rm sc} = \frac{3}{5}\left(\frac{3\pi}{2}\right)^{2/3} = 1.68647 \; . \tag{4.19}$$

  This critical threshold for (spherical) collapse plays a central role in models that use statistics of the initial density fluctuation field to describe the clustering of virialized objects.

- Linear theory furnishes the clock of spherical collapse. You just have to linearly extrapolated any initial density fluctuation $\delta_i(t_i)$ (now assumed to be a pure growing mode):

$$\delta_{\rm L}(t) = \delta_i \left(\frac{t}{t_i}\right)^{2/3} \; . \tag{4.20}$$

  It collapses and virializes at a time $t_{\rm vir}$ given by $\delta_{\rm L}(t_{\rm vir}) \equiv \delta_{\rm sc}$.

4.2 Press-Schechter and excursion set theory

The fundamental difficulty at predicting the number density of galaxies is that the galactic scales have already gone nonlinear. The average density in a spherical region of radius R indeed is

$$\bar{\rho}_m = \frac{3M}{4\pi R^3} \equiv \Omega_m \rho_c(t_0) \; . \tag{4.21}$$

On inverting this relation we find

$$R = 0.951\; h^{-1}\text{Mpc} \left(\frac{M}{10^{12}\Omega_m\; \text{M}_\odot/h}\right)^{1/3} \; . \tag{4.22}$$

A typical galaxy has a mass $M = 10^{12}\; \text{M}_\odot/h$ and, therefore, roughly corresponds to fluctuations on scales of order $k \sim 1\; h\text{Mpc}^{-1}$ well into the nonlinear regime.

However, the spherical collapse model, when combined with a statistical description of the linear density fluctuations, gives, e.g., dark matter halo abundances which are in good agreement with numerical simulations. These can be populated using a simple HOD (halo occupation distribution) to predict the abundance of galaxies.

4.2.1 Filling factor

The analysis begins with the assumption that the initial density fluctuation (and, thus, the primordial curvature perturbations) is a Gaussian random field. In other words, the phases of the Fourier modes $\delta(\mathbf{k})$ are random while their amplitude follow a Rayleigh distribution.

Since the linear theory prediction is the clock for the spherical collapse, it is convenient to work with the initial density linearly extrapolated to the redshift z at which virialized objects are observed. This field does not need to obey the physical constraint $\delta > -1$ because it is merely the linear extrapolation of a density contrast of much smaller magnitude.

In the spirit of the spherical collapse, we further smooth this field with a tophat filter of radius $R = (3M/4\pi)^{1/3}$ to define perturbations on a mass scale M. We thus have

$$\delta_M(\mathbf{x}, z) \equiv \int \frac{d^3k}{(2\pi)^3}\, \delta(\mathbf{k}, z)\hat{W}_\text{T}(k; R)e^{i\mathbf{k}\cdot\mathbf{r}} \tag{4.23}$$

at the comoving *Lagrangian* (i.e. initial) position $\mathbf{x}$ (so that the position of each mass element is independent of time). The Fourier modes of the linearly extrapolated initial density field are given by

$$\delta(\mathbf{k}, z) = \mathcal{M}(k, z)\Psi(\mathbf{k}, z_i) \; , \tag{4.24}$$

where

$$\mathcal{M}(k, z) = \frac{2}{3}\frac{k^2 T(k) D(z)}{\Omega_m H_0^2} \tag{4.25}$$

maps the gravitational potential $\Psi(\mathbf{k}, z_i)$ deep in matter domination onto fluctuations in the linear density field at redshift z.

We assume that fluctuations $\delta_M(\mathbf{x}, z)$ above the critical threshold δ_sc will correspond to virialized objects of mass greater than M at redshift z. Since the (1-point) distribution of linear overdensities $\delta_M(\mathbf{x}, z)$ satisfies

$$P(\delta_M)d\delta_M = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{\delta_M^2}{2\sigma^2}}\, d\delta_M \; , \tag{4.26}$$

where $\sigma \equiv \sigma(M,z)$ is the rms variance of δ_M, the fraction of Lagrangian volume in halos of mass greater than M, or *filling factor*, is

$$F(>M) = \frac{1}{\sqrt{2\pi}\sigma}\int_{\delta_{\rm sc}}^{\infty} d\delta\, e^{-\frac{\delta_M^2}{2\sigma^2}} = \frac{1}{2}\,{\rm erfc}\left(\frac{\nu}{\sqrt{2}}\right) \;, \tag{4.27}$$

where $\nu \equiv \delta_{\rm sc}/\sigma(M,z)$ is the peak height or significance and erfc(x) is the complementary error function. We can define a characteristic mass $M_\star(z)$ at each redshift z through the requirement $\sigma(M_\star(z),z) = \delta_{\rm sc}$. This is the typical mass of halos collapsing at redshift z.

Assuming that all the mass is in virialized objects, $F(>0)$ should be equal to unity. For the hierarchical power spectra we consider, $\sigma(M,z)$ becomes arbitrarily large as M decreases. However, erfc(0)=1 so that only half of the mass in the Universe is actually contained in dark matter halos. Press & Schechter already noticed that this problem arises because Eq.(4.27) only counts overdense regions. They argued that underdense regions will collapse onto overdense regions and multiply $F(>M)$ by a factor of two. Though the sense of this effect is certainly such that more mass will be contained in bound objects, it is unclear that it should precisely lead to a factor of two. We will come back to this point shorlty. Meanwhile, we shall introduce an ad hoc multiplicative factor of two.

4.2.2 The halo mass function

Differentiating with respect to the mass M yields the filling factor of virialized objects at redshift z in the range $[M, M+dM]$. On multiplying by $\bar{\rho}_m/M$, we thus obtain the differential number density $\bar{n}(M)$ of virialized objects per unit of mass and volume or, equivalently, the *halo mass function*. Including the "fudge factor" of 2, this is

$$\frac{dn}{dM} \equiv \bar{n}(M) = \frac{\bar{\rho}_m}{M}\left|\frac{dF(>M)}{dM}\right| = \sqrt{\frac{2}{\pi}}\frac{\bar{\rho}_m}{M^2}\,\nu e^{-\nu^2/2}\frac{d\ln\nu}{d\ln M} \;. \tag{4.28}$$

For the favoured CDM power spectrum, $d{\rm ln}\nu/d{\rm ln}M \approx 0.1 - 0.3$ in the range $M \sim 10^{10} - 10^{16}\ {\rm M}_\odot/h$. With a comoving matter density given by $\bar{\rho}_m = 27.75\times 10^{10}\,\Omega_m\ {\rm h^2 M_\odot Mpc^{-3}}$, the abundance of halo per logarithmic mass interval is approximately $dn/d\ln M \sim 3\,(M/10^{10}\ {\rm M}_\odot/h)^{-1}\ {\rm h^3Mpc^{-3}}$ at the characteristic mass $M_\star(z)$. The high mass tail is exponentially suppressed above $M_\star(z)$ while at low mass, the halo number density diverges as $\bar{n}(M) \propto M^{-2}$.

To get a better impression of the Press-Schechter mass function, we consider the special case of a powerlaw spectrum $P_\delta(k) = k^n$ for the linear mass density field δ. Clearly, this is a good description of the linear density power spectrum only far above or below its maximum (see exercises). However, the typical scales that virialize today ($k \gtrsim 1\ h^{-1}{\rm Mpc}$) are relatively small and, therefore, a powerlaw with $-1 \lesssim n \lesssim -2$ should be a valid approximation. The variance in tophat perturbations of mass M is approximately

$$\sigma^2(M) \sim \int_0^{1/R} dk\, k^{(n+2)} = AM^{-\gamma} \;, \tag{4.29}$$

where $\gamma = 1 + n/3 \sim 0.5$ and A is proportional to the normalization of the primordial power spectrum. Therefore, the characteristic mass is $M_\star = (A/\delta_{\rm sc}^2)^{1/\gamma}$. Furthermore, since $\sigma^2(M) \propto t^{4/3}$ in the matter dominated epoch, we have $A \propto t^{4/3}$, so that

$$M_\star(t) \propto t^{4/3\gamma} \;. \tag{4.30}$$

This shows that, in CDM models, low-mass structures (individual galaxies) form at early times, whereas large mass fluctuations (such as galaxy clusters) collapse only later; i.e. structure formation is *hierarchical* or *bottom-up*. Writing the peak height as $\nu = (M/M_\star)^{\gamma/2}$, the halo mass function eventually takes the form

$$\bar{n}(M) = \frac{1}{\sqrt{2\pi}} \frac{\gamma \bar{\rho}_m}{M^2} \left(\frac{M}{M_\star}\right)^{\gamma/2} \exp\left[-\frac{1}{2}\left(\frac{M}{M_\star}\right)^{\gamma}\right] , \tag{4.31}$$

in which all the time dependence has been absorbed into the variation of $M_\star$ with cosmic time.

- The Press-Schechter approach does not account for the possibility that $\delta_M(\mathbf{x}, z)$ may be less than $\delta_{\rm sc}$ on scale M, yet $\delta_{M'}(\mathbf{x}, z)$ may be larger than $\delta_{\rm sc}$ for some $M' > M$. It seems natural that the perturbation on the larger scale M' will collapse and form a virialized objects, overwhelming the more diffuse patches within it. Accounting for this so-called *cloud-in-cloud problem* will increase the fraction of mass in collapsed objects and indeed expain the missing factor of two in the Press-Schechter filling fraction $F(> M)$
- To solve the cloud-in-cloud problem, it is necessary to compute the largest mass scale M (i.e. the largest smoothing scale R) at which the critical threshold $\delta_{\rm sc}$ is exceeded. Let us pick up some random comoving Lagrangian position $\mathbf{x}$ and denote $S = \sigma^2(R)$. Since S is a monotonically decreasing function of the smoothing scale R, we can also use it to label the smoothing scale. Therefore, $\delta(S)$ will designate the trajectory described by the linear density contrast at that Lagrangian position $\mathbf{x}$ when the amount of smoothing varies.

Analytic predictions for the halo mass function $\bar{n}(M)$ can be compared to the outcome of N-body simulations, which follow the motion of dark matter "particles" (these typically have a mass 10^{10} $M_\odot/h$) in a large volume of the Universe. For the sake of comparison, it is convenient to express $\bar{n}(M)$ in terms of the multiplicity function $f \sim \mathcal{F}/\nu$,

$$\bar{n}(M) = \frac{\bar{\rho}_m}{M^2} \nu f(\nu, \dots) \frac{d \ln \nu}{d \ln M} . \tag{4.32}$$

In principle, f could depends on variables other than the peak significance. However, when f depends solely on ν, the halo mass function can be scaled to a universal (self-similar) form which is independent of cosmology, redshift and power spectrum. This is the case of the Press-Schechter mass function for instance, for which

$$\nu f(\nu) = \sqrt{\frac{2}{\pi}}\, \nu\, e^{-\nu^2/2} . \tag{4.33}$$

While the Press-Schechter prediction broadly agrees with measurements from numerical simulations, the agreement is far from perfect. Various improvements have been considered since then, including ellipsoidal collapse and excursion set theory.