Chapter 13

Relativistic cosmology I: general geometry.

13.1 A continuous medium as a model of the Universe.

When considering the Universe as a whole, one assumes that it is filled with a continuous medium (fluid or gas), whose state can be described by scalar fields such as mass density or pressure, vector fields such as the velocity of flow, or tensor fields, e.g. an electromagnetic field. This is a rather crude approximation, since our real Universe has a "granular" structure. Its basic units are stars, and the relevant information is the number of stars in a given volume rather than the average mass density in that volume. Moreover, the view on which objects should be "elementary cells" of the cosmic fluid has been changing with time. In the times of Hubble (1920s and 1930s), these were the galaxies. In later times, the galaxy clusters took over. In still later years, it was found that galaxies and galaxy clusters tend to occupy edges of large volumes of space that are almost empty (called voids, see Sec. 15.5). The current view is that the elementary units of the Universe are groups of voids. These changes in the definition of the elementary unit of the Universe were adopted in order to save the assumption of homogeneity and isotropy of the Universe "in the large".

This assumption deserves a separate comment. Astronomical observations provide reliable quantitative information only about a small neighbourhood of the Solar System. With increasing distance from the Earth, uncertainties enter. In order to describe the Universe as a whole, the results of local observations have to be extrapolated to large volumes, and then the conclusions from the extrapolation have to be tested observationally. The extrapolations rely on assumed models. Hence, if a given extrapolation leads to a correct prediction, this does not mean that it was the only one possible.

The widest extrapolation is contained in the so-called **cosmological principle**. It stems from the idea of Copernicus who noted that the Earth is not at the centre of the world, but occupies an unimportant position in the Solar System. Afterwards, the Earth had been "degraded" a few more times: it turned out that even the Sun is not at the centre of the Universe, but is one of a great number of stars that are similar to each other, and that our Galaxy is also one of many, not the greatest one and not placed at any special position. The cosmological principle is a summary of this line of thinking. In its weaker form, it says: the Earth occupies a position in the Universe that is not in any way preferred. But, quite often, it is expressed in the extreme form: all positions in the Universe are equivalent, the results of observations do not depend on the observer's position in the Universe. Whether one believes in the cosmological principle or not, and in which version of it, it must be remembered that this principle is not a summary of observational results, but an *assumption*, upon which the theory of the structure of the Universe is built. It was an acceptable working hypothesis when constructing the first-ever models of the Universe in the 1920s [38, 108] because at that time no observational data contradicted it. Today, the cosmological principle still has no direct observational verification,[31] while models not obeying it are known, see chapter 15. The fact that the whole of observational cosmology is based on the Friedmann–Lemaître models is a consequence of mental inertia. Natural sciences are said to use the criterion of consistency of theory with observations/experiments. So, in order to verify the cosmological principle, alternatives to it have to be considered and compared with observations. Working always from within the same theory, we do not expose its basic assumptions to tests.

Observations show that the space is *approximately* isotropic around us.[32] By the cosmological principle, the space should thus be isotropic around every other point. A space that is isotropic around every point is homogeneous. This argument points to the Robertson – Walker spacetimes introduced in Sec. 9.8.

Before we come to the R–W models, we shall discuss the fluid model of the Universe in the background of a general geometry. We assume that each point in the Universe can be assigned energy density, pressure and the four-vector of velocity of the fluid particle that passes through it. We also assume that the cosmic matter treated in this way obeys the equations of hydrodynamics known from laboratory – which is another bold assumption.

13.2 Optical observations in the Universe – part I

13.2.1 The geometric optics approximation

We will now consider the influence of the geometry of the Universe and of the kinematics of the cosmic medium on the propagation of light-rays. The approach presented below is based on the papers by Ellis [102, 109] and Kristian and Sachs [110].

[108] G. Lemaître, *Ann. Soc. Sci. Bruxelles* **A47**, 19 (1927); English translation (somewhat updated): *Mon. Not. Roy. Astr. Soc.* **91**, 483 (1927). Faithful translation: *Gen. Relativ. Gravit.* **45**, 1635 (2013), with an editorial note by J.-P. Luminet,*Gen. Relativ. Gravit.* **45**, 1619 (2013).

[31] Because of the difficulties in determining the distances to other galaxies.

[32] At present, the main argument supporting this statement is the isotropy of the microwave background radiation, but the statement had been made long before the CMB radiation was discovered.

[109] G. F. R. Ellis, in: *Cargése Lectures in Physics*, vol. 6. Edited by E. Schatzman. Gordon and Breach, New York 1973, p. 1 (1973).

[110] J. Kristian and R. K. Sachs, *Astrophys. J.* **143**, 379 (1966); reprinted in: *Gen. Relativ. Gravit.* **43**, 337 (2011), with an editorial note by G. F. R. Ellis, *Gen. Relativ. Gravit.* **43**, 331 (2011) and Kristian's biography by A. Krasiński, *Gen. Relativ. Gravit.* **43**, 335 (2011).

We assume that the light ray avoids large-scale accumulations of charges or currents, and that the electromagnetic field in it is weak enough not to influence the geometry of the spacetime (i.e. it is a *test field*). Then, the Maxwell equations read

$$F_{[\alpha\beta;\gamma]} \equiv F_{[\alpha\beta,\gamma]} = 0, \quad (13.1)$$
$$F^{\alpha\beta}{}_{;\beta} = 0. \quad (13.2)$$

We will seek solutions of this set in the form of waves:

$$F_{\alpha\beta} = G_{\alpha\beta} \sin(S + \phi), \quad (13.3)$$

where the amplitude $G_{\alpha\beta}$ is assumed to vary slowly compared to the phase S; ϕ is a constant. Thus, we assume that each differentiation of $G_{\alpha\beta}$ introduces a factor ε, which is a small parameter. If $G_{\alpha\beta}$ can be developed in a power series with respect to ε, then

$$G_{\alpha\beta} = B^0_{\alpha\beta} + \sum_{i=1}^{\infty} \varepsilon^i B^i_{\alpha\beta}. \quad (13.4)$$

We now assume that the Maxwell equations are fulfilled at each order in ε, and substitute (13.4) in (13.1) – (13.2). Denoting[33]

$$S_{,\alpha} = k_\alpha \quad (13.5)$$

we obtain at the zeroth order:

$$(B^0)^{\alpha\beta} k_\beta = 0, \qquad k_{[\alpha} B^0_{\beta\gamma]} = 0. \quad (13.6)$$

and at the first order

$$(B^0)^{\alpha\beta}{}_{;\beta} \sin(S + \phi) = -\varepsilon (B^1)^{\alpha\beta} k_\beta \cos(S + \phi), \quad (13.7)$$
$$B^0_{[\alpha\beta;\gamma]} \sin(S + \phi) = -\varepsilon B^1_{[\alpha\beta} k_{\gamma]} \cos(S + \phi). \quad (13.8)$$

Since $(B^i)^{\alpha\beta} = (B^i)^{[\alpha\beta]}$, for all i, we have $(B^0)^{\alpha\beta}{}_{;\beta} = \left[\sqrt{-g}(B^0)^{\alpha\beta}\right]{}_{,\beta} / \sqrt{-g}$, so this term is of order ε by virtue of our assumptions. The same applies to (13.8), where $B^0_{[\alpha\beta;\gamma]} = B^0_{[\alpha\beta,\gamma]}$. Equations (13.7) – (13.8) show that the first-order terms act as sources in the Maxwell equations for the zero-th order terms. Thus, in a curved space, the electromagnetic wave does not in fact propagate into vacuum – the higher-order terms act as a medium with currents and charges, on which the zero-th order wave may be dispersed. Similarly, the first order terms are influenced by second-order terms, and so on. (In a flat space in Cartesian coordinates, a constant $B^0_{\alpha\beta}$, with all $B^i_{\alpha\beta} \equiv 0$ for $i \geq 1$ is a solution of (13.7) – (13.8).) Assuming the scheme is self-consistent and the "tail" terms remain small, we will now consider the consequences of (13.6). The second of (13.6) can be written as

$$k_\alpha B^0_{\beta\gamma} + k_\beta B^0_{\gamma\alpha} + k_\gamma B^0_{\alpha\beta} = 0. \quad (13.9)$$

Contracting this with k^α and making use of (13.6) we immediately obtain

$$k^\alpha k_\alpha = 0, \quad (13.10)$$

[33] With (13.5) fulfilled, the *rotation* of k_α, defined later in this section, is zero. Thus, the geometric optics approach is more general: the wave description does not allow rotating congruences of rays.

$$\Longrightarrow k_{\alpha;\beta}k^{\alpha} \;=\; 0. \tag{13.11}$$

But since $k_\alpha = S,_\alpha$, we have $k_{\alpha;\beta} = k_{\beta;\alpha}$, and then from (13.11)

$$k_{\beta;\alpha}k^{\alpha} = 0, \tag{13.12}$$

so k^α is a null geodesic vector, and its parametrisation is affine.

Contracting (13.9) with $(B^0)^{\beta\gamma}$ and using (13.6) we obtain

$$B^0_{\alpha\beta}(B^0)^{\alpha\beta} = 0. \tag{13.13}$$

Contracting (13.9) with $(B^0)^{\delta\gamma}$ and using (13.6) once more we obtain

$$k_\alpha B^0_{\beta\gamma}(B^0)^{\delta\gamma} - k_\beta B^0_{\alpha\gamma}(B^0)^{\delta\gamma} = 0. \tag{13.14}$$

This means that $B^0_{\alpha\gamma}(B^0)^{\delta\gamma}$ with any value of δ is proportional to k_α, thus

$$B^0_{\alpha\gamma}(B^0)^{\delta\gamma} = U^\delta k_\alpha, \tag{13.15}$$

where U^δ is the (vectorial) proportionality factor. With δ lowered, the left-hand side of (13.15) is symmetric in $(\alpha\delta)$, so $U_\delta k_\alpha - U_\alpha k_\delta = 0$ and $U_\alpha = \mu k_\alpha$. Thus, finally

$$B^0_{\alpha\gamma}(B^0)_\delta{}^\gamma = \mu k_\alpha k_\delta. \tag{13.16}$$

The energy-momentum tensor of a general electromagnetic field is [11]

$$T^{\alpha\beta} = \frac{1}{4\pi}\left(F^{\alpha\mu}F_\mu{}^\beta + \frac{1}{4}g^{\alpha\beta}F_{\mu\nu}F^{\mu\nu}\right). \tag{13.17}$$

From (13.13) and (13.16) we see that the tensor (13.17) for the field (13.3) – (13.4) is

$$T_{\alpha\beta} = \frac{1}{4\pi}\ \left(\mu k_\alpha k_\beta + O(\varepsilon)\right), \tag{13.18}$$

which is, up to terms linear in ε, a perfect-fluid-type medium with the "four-velocity" k_α. With k_α being null, the velocity of flow equals the velocity of light, i.e., (13.18) corresponds to a stream of photons. From $T^{\alpha\beta};_\beta = 0$, making use of (13.12), we obtain

$$(\mu k^\alpha)\,;_\alpha = 0. \tag{13.19}$$

13.2.2 The redshift

An observer moving with the four-velocity u^α will measure the rate of change of phase of the light wave $v_p = S,_\alpha u^\alpha = k_\alpha u^\alpha$. Within a short time-interval Δs, the phase will thus change by $\Delta S = k_\alpha u^\alpha \Delta s$. For another observer, moving with the velocity u^α_1 and measuring the change of phase at another spacetime point, where $k^\alpha = k^\alpha_1$, the same change of phase ΔS will in general take a different time-interval, Δs_1: $\Delta S = k_{1\alpha}u^\alpha_1 \Delta s_1$. Hence:

$$\frac{\Delta s_1}{\Delta s_2} = \frac{(k_\alpha u^\alpha)_2}{(k_\alpha u^\alpha)_1}. \tag{13.20}$$

If the electromagnetic wave is periodic with frequency ν, then $\Delta S = 2\pi\nu\Delta s$. Hence, for the same ΔS measured by two observers we have $\nu_1\Delta s_1 = \nu_2\Delta s_2$, so

$$\frac{\nu_2}{\nu_1} = \frac{\Delta s_1}{\Delta s_2} = \frac{(k_\alpha u^\alpha)_2}{(k_\alpha u^\alpha)_1}. \tag{13.21}$$

This is the formula for the cosmological **redshift** that applies in all cosmological models. Let the subscripts e and o denote quantities calculated at the point of emission of the light ray and at the point of detection, respectively. Then, from the definition of redshift:

$$z = \frac{\lambda_o - \lambda_e}{\lambda_e} = \frac{\lambda_o}{\lambda_e} - 1. \tag{13.22}$$

But $\lambda_o/\lambda_e = \nu_e/\nu_o$ (because $c = \lambda\nu$ is constant), hence:

$$1 + z = \frac{\nu_e}{\nu_o} = \frac{(k_\alpha u^\alpha)_e}{(k_\alpha u^\alpha)_o}. \tag{13.23}$$

The direction of the ray at the observer is determined by the unit spacelike vector n^α:

$$n^\alpha = \xi \left(\delta^\alpha{}_\beta - u^\alpha u_\beta\right) k^\beta, \qquad n_\alpha n^\alpha = -1, \tag{13.24}$$

which is collinear with the projection of k^α on the hypersurface of constant time of the observer. Substituting the first of (13.24) in the second we obtain $\xi^2 = \left(k_\rho u^\rho\right)^{-2}$. Hence,

$$n^\alpha = -\frac{1}{k_\rho u^\rho}\, k^\alpha + u^\alpha \qquad (\Longrightarrow n^\alpha u_\alpha = 0)\,. \tag{13.25}$$

The minus sign is there because n^α points *towards* the source of light, opposite to the direction of k^α at the observer.

So far, our considerations applied in every cosmological model and up to an arbitrary distance; the only approximation was that connected with geometric optics. In order to apply (13.23) to results of observations, one has to integrate the equations of a null geodesic, which is always difficult. But when $z \ll 1$, an approximate version of (13.23) may be used. Then, (13.22) becomes $z = \mathrm{d}\lambda/\lambda$, and in (13.23) we have

$$z = \frac{(k_\alpha u^\alpha)_e - (k_\alpha u^\alpha)_o}{(k_\alpha u^\alpha)_o} = \frac{\mathrm{d}\,(k_\alpha u^\alpha)}{(k_\alpha u^\alpha)_o}. \tag{13.26}$$

The change $\mathrm{d}(\cdot)$ in (13.26) is along the light ray. Let v be the affine parameter on the ray, then

$$\mathrm{d}\,(k_\alpha u^\alpha) = D\,(k_\alpha u^\alpha) = \left(k_{\alpha;\beta} k^\beta u^\alpha\right)_o \mathrm{d}v + \left(k_\alpha u^\alpha{}_{;\beta}\, k^\beta\right)_o \mathrm{d}v. \tag{13.27}$$

The first term vanishes in consequence of (13.12), and in the second term we substitute the decomposition (12.28) and use the antisymmetry of $\omega_{\alpha\beta}$. The result is

$$\mathrm{d}\,(k_\alpha u^\alpha) = \left(u_{\alpha;\beta} k^\alpha k^\beta\right)_o \mathrm{d}v = \left[\left(\sigma_{\alpha\beta} k^\alpha k^\beta\right)_o - \frac{1}{3}\theta (k_\alpha u^\alpha)_o{}^2 + (k_\alpha \dot{u}^\alpha)_o (k_\alpha u^\alpha)_o\right] \mathrm{d}v. \tag{13.28}$$

Using (13.25) and (12.33) we get

$$\mathrm{d}\,(k_\alpha u^\alpha) = \left(k_\rho u^\rho\right)_o{}^2 \left(\sigma_{\alpha\beta} n^\alpha n^\beta - \frac{1}{3}\theta - n_\alpha \dot{u}^\alpha\right)_o \mathrm{d}v. \tag{13.29}$$

Note in addition that (13.25) and (13.10) imply

$$k^\rho n_\rho = k^\rho u_\rho, \tag{13.30}$$

and that $\left[-\left(k^\rho n_\rho\right)_o \mathrm{d}v\right]$ is the distance in the rest space of the observer travelled by the front of the light wave corresponding to the change $\mathrm{d}v$ in the affine parameter, i.e. it is the distance $\delta\ell$ between the light-source and the observer. Using (13.29) in (13.26) we thus get

$$z = \left(-\sigma_{\alpha\beta} n^\alpha n^\beta + \frac{1}{3}\theta + n_\alpha \dot{u}^\alpha\right)_o \delta\ell. \tag{13.31}$$

This shows that for $z \ll 1$ rotation has no influence on z while $\sigma_{\alpha\beta}$ and $\dot{u}^\alpha$ introduce anisotropy in z. Equation (13.31) is still correct in every cosmological model, but only for light-sources that are near to the observer. Its advantage is that in using it one does not need to integrate any differential equations.

13.3 The optical tensors.

We shall now apply a similar reasoning to that in Sec. 12.2 to families of null curves. A space locally orthogonal to a null vector k contains k, so now the projection has to be defined in a different way. Given k^α, we first define a second null vector ℓ^α that is tangent to the same light cone and obeys:

$$\ell^\alpha k_\alpha = 1, \qquad \ell^\alpha \ell_\alpha = 0. \tag{13.32}$$

Note that ℓ^α is not defined uniquely. If m^α is an arbitrary spacelike vector of unit length ($m^\alpha m_\alpha = -1$) orthogonal to both k^α and ℓ^α, then $\ell'^\alpha = \ell^\alpha + \frac{1}{2}b^2 k^\alpha + b m^\alpha$ obeys (13.32) as well, where b is an arbitrary parameter.

Then we define the projection tensor on the surface (this time 2-dimensional) that is orthogonal to both ℓ^α and k^α:

$$p_{\alpha\beta} = g_{\alpha\beta} - \ell_\alpha k_\beta - k_\alpha \ell_\beta \Longrightarrow p_{\alpha\beta} k^\beta = p_{\alpha\beta}\ell^\beta = 0. \tag{13.33}$$

Tangent planes to this surface do not include ℓ^α or k^α – because a vector that is orthogonal to two linearly independent null vectors must be spacelike, see Exercise 1.

Now assume that k^α is a null and geodesic, affinely parametrised vector *field*. Then

$$k^\mu k_\mu = 0 = k^\mu k_{\alpha;\mu}. \tag{13.34}$$

We define

$$A_{\alpha\beta} \overset{\text{def}}{=} k_{\rho;\sigma} p^\rho{}_\alpha p^\sigma{}_\beta. \tag{13.35}$$

Then the following holds

$$k_{\alpha;\beta} = A_{\alpha\beta} + a_\alpha k_\beta + k_\alpha b_\beta, \tag{13.36}$$

where

$$a_\alpha = \ell^\rho k_{\alpha;\rho} - \frac{1}{2} k_{\rho;\sigma}\ell^\rho \ell^\sigma k_\alpha, \qquad b_\alpha = \ell^\rho k_{\rho;\alpha} - \frac{1}{2} k_{\rho;\sigma}\ell^\rho \ell^\sigma k_\alpha. \tag{13.37}$$

It follows that $a_\alpha k^\alpha = b_\alpha k^\alpha = 0$. Now we apply to $A_{\alpha\beta}$ the decomposition into the trace, the trace-free symmetric part and the antisymmetric part, similar to the one in Sec. 12.2:

$$A_{\alpha\beta} = \omega_{\alpha\beta} + \sigma_{\alpha\beta} + p_{\alpha\beta}\theta \tag{13.38}$$

(by tradition, the last term does not include the coefficient 1/2), where

$$\omega_{\alpha\beta} \overset{\text{def}}{=} A_{[\alpha\beta]} \tag{13.39}$$

is called the **rotation** of the family of null curves,

$$\theta \overset{\text{def}}{=} \frac{1}{2} g^{\alpha\beta} A_{\alpha\beta} \tag{13.40}$$

is called the **expansion** of the family, and

$$\sigma_{\alpha\beta} \overset{\text{def}}{=} A_{(\alpha\beta)} - p_{\alpha\beta}\theta \tag{13.41}$$

is called the **shear** of the family. The geometric interpretation of rotation, expansion and shear is similar to that in hydrodynamics; this time the respective changes apply to images of an object projected by the family of light rays onto 2-surfaces orthogonal to the family. Like in hydrodynamics, rotation and shear do or do not vanish only simultaneously with the corresponding scalars, defined below. Also, the following equations are useful:

$$\begin{aligned} p^\alpha{}_\rho p^\rho{}_\beta &= p^\alpha{}_\beta, \qquad g^{\alpha\beta} p_{\alpha\beta} = p^{\alpha\beta} p_{\alpha\beta} = 2 \\ p^{\alpha\beta}\sigma_{\alpha\beta} &= g^{\alpha\beta}\sigma_{\alpha\beta} = 0, \\ k^\beta \omega_{\alpha\beta} &= k^\beta \sigma_{\alpha\beta} = \ell^\beta \omega_{\alpha\beta} = \ell^\beta \sigma_{\alpha\beta} = 0. \end{aligned} \tag{13.42}$$

The scalars of rotation, expansion and shear are then

$$\omega^2 \overset{\text{def}}{=} \frac{1}{2}\omega_{\alpha\beta}\omega^{\alpha\beta} = \frac{1}{2} k_{[\alpha;\beta]} k^{\alpha;\beta}, \tag{13.43}$$

$$\theta = \frac{1}{2} k^\mu{}_{;\mu}\, , \tag{13.44}$$

$$\sigma^2 \overset{\text{def}}{=} \frac{1}{2}\sigma_{\alpha\beta}\sigma^{\alpha\beta} = \frac{1}{2} k_{(\alpha;\beta)} k^{\alpha;\beta} - \theta^2. \tag{13.45}$$

As seen, these quantities depend only on k^α, not on the auxiliary field ℓ^α.

Take the Ricci identity for the null field k^α:

$$k_{\alpha;\beta\gamma} - k_{\alpha;\gamma\beta} = R_{\rho\alpha\beta\gamma} k^\rho. \tag{13.46}$$

The following equations will be useful in further calculation:

$$g^{\alpha\beta} k_{\alpha;\beta\gamma} \equiv \left(k^\beta{}_{;\beta}\right){}_{;\gamma} = 2\theta_{,\gamma}\, , \tag{13.47}$$

$$k^\gamma g^{\alpha\beta} k_{\alpha;\gamma\beta} \equiv k^\gamma k^\beta{}_{;\gamma\beta} = -k^{\gamma;\beta} k_{\beta;\gamma}, \tag{13.48}$$

$$k^{\gamma;\beta} k_{\beta;\gamma} = 2\left(\sigma^2 - \omega^2 + \theta^2\right). \tag{13.49}$$

Now contract (13.46) with $k^\gamma g^{\alpha\beta}$ and use (13.47) – (13.49) The result is the equation of evolution of θ, analogous to the Raychaudhuri equation (12.44):

$$k^\gamma \theta_{,\gamma} + \sigma^2 - \omega^2 + \theta^2 = -\frac{1}{2} R_{\rho\gamma} k^\rho k^\gamma. \tag{13.50}$$

13.4 The apparent horizon.

The event horizon, which we defined for the Schwarzschild metric in Sec. 11.8, could be determined observationally only if we knew the whole future evolution of the spacetime. It is more realistic to use only such notions that can be identified by local observations of limited duration. The apparent horizon is one.

In a flat spacetime, if a flash of light is sent from both sides of a closed surface, then the light rays converge inside the surface and diverge outside. The convergence/divergence is measured by the scalar of expansion (13.44). However, in the vicinity of a singularity the *outward-directed* bundle of rays is convergent. This is easy to see in the Kruskal diagram (Fig. 11.9): under the horizon H_1 r is decreasing along every light ray. Hence, the area of the light front is decreasing for both the outward- and inward-bundle. A closed surface from which it is impossible to send a diverging bundle of light rays is called a **closed trapped surface**. Then, an **apparent horizon** is the outer envelope of the region in which closed trapped surfaces exist.

There are two kinds of trapped surfaces and apparent horizons, as seen in Fig. 11.9: the **future-trapped surfaces** and **future apparent horizons** in region II, and the **past-trapped surfaces** and **past apparent horizons** in region IV. For a past-trapped surface, the light rays converge towards the past. In other words, for a past-trapped surface S_p, both *ingoing* bundles of rays (i.e. those that simultaneously reach S_p both from inside and from outside) are necessarily *diverging*, while for a future-trapped surface S_f, both *outgoing* bundles of rays are necessarily converging (where "outgoing" means starting their journey simultaneously at S_f, both inward and outward).

In the Schwarzschild spacetime, the apparent horizons coincide with the event horizons, but in nonstatic spacetimes they are different. Examples will be given in Sec. 15.8.

13.5 Optical observations in the Universe – part II

13.5.1 The area distance

In curved spacetime, it is a problem to define the distance between two objects. Abstract geometric definitions, such as the integral of ds along a uniquely defined spacelike path, have no relation to astronomical practice. We need a definition that would yield a quantity measurable by means of observations. The reasoning used here is adapted from Ref. [102],[34] and the definitions of distance go back to Etherington[111]. The subscript 'e' will denote quantities calculated at the emitter, 'o' will denote quantities at the observer.

An observer moving with a 4-velocity u_o^{α} would measure the flux of radiation reaching her (i.e. the energy flowing through a unit surface area in a unit of time) to be equal to

[34] The author of Ref. [102] used units in which the velocity of light $c = 1$. This had the undesirable consequence that the energy density of radiation and its flux became equal.

[111] I. M. H. Etherington, *Phil. Mag., ser. 7* **15**, 761 (1933), reprinted in *Gen. Relativ. Gravit.* **39**, 1055 (2007) with an editorial note by G. F.R. Ellis, *Gen. Relativ. Gravit.* **39**, 1047 (2007) and author's biography by A. Krasiński, *Gen. Relativ. Gravit.* **39**, 1053 (2007).

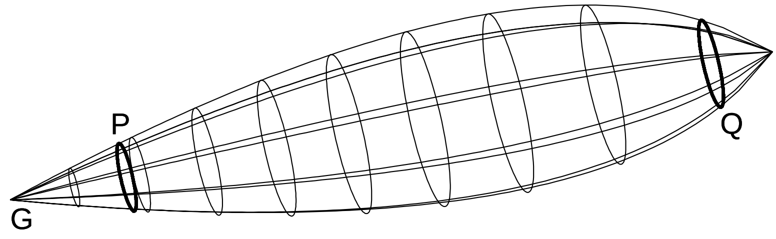

Figure 13.1: A non-rotating null congruence in vacuum or in a perfect fluid necessarily has its expansion scalar decreasing along the rays. This may (if the distance is sufficiently large) eventually cause the expansion to become negative, which results in refocussing. The observer at Q would see the source G to be anomalously bright and large, if she were able to cancel the redshift. Actually, when shear is nonzero, the geodesics will be refocussed not to a point, but to a caustic surface.

$\mathcal{F}_o = cT_{\alpha\beta}u_o^\alpha u_o^\beta$, where $T_{\alpha\beta}$ is the energy-momentum tensor of radiation, given by (13.15), calculated at the observer's position. Thus, neglecting terms of order ε:

$$\mathcal{F}_o = \frac{c\mu_o}{4\pi}\ \left(k_\alpha u^\alpha\right)_o^2 . \tag{13.51}$$

On the other hand, combining the conservation equation (13.19) with the surface-area-propagation equation $k^\mu;_\mu = \mathrm{d}(\ln S_n)/\mathrm{d}s$ (see Exercise 2) we see that $\mu\delta S$ is constant along null geodesics, thus

$$(\mu\delta S)|_{s=s_1} = (\mu\delta S)|_{s=s_2} \tag{13.52}$$

(δS is the area of an orthogonal cross-section of the bundle of rays). From (13.51) and (13.52) we obtain $\mathcal{F}_o = \text{constant}\left(k_\alpha u^\alpha\right)_o^2/(\delta S)_o$. Remembering that $\left(k_\alpha u^\alpha\right)_e$ is constant along each null geodesic and using (13.20), we obtain

$$\mathcal{F}_o = \frac{C_\mathcal{F}}{\delta S_o}\ \left[\frac{\left(k_\alpha u^\alpha\right)_o}{\left(k_\alpha u^\alpha\right)_e}\right]^2 = \frac{C_\mathcal{F}}{\delta S_o\left(1+z_o\right)^2}, \tag{13.53}$$

where $C_\mathcal{F}$ is constant in each bundle (it depends on the solid angle occupied by the bundle). Thus, $\mathcal{F}_o$ decreases with increasing redshift. However, δS_o does not necessarily increase with distance, as can be seen from the "null Raychaudhuri" equation (13.50). Assume that $\omega = 0$ and $R_{\mu\nu} = 0$ (vacuum) for simplicity. Then (13.50) shows that $\theta,_\mu k^\mu < 0$, that is, θ decreases along the ray and may become negative. This remains true in a perfect fluid or dust, where $R_{\alpha\beta}k^\alpha k^\beta = \kappa(\epsilon + p)\left(k_\alpha u^\alpha\right)^2 > 0$. Consequently, *curvature causes null geodesics to converge*, even in vacuum. Thus, a situation shown in Fig. 13.1 is possible: an observer at Q will see the light source at G to have the same brightness as seen by the observer at P. For Q, the source will appear anomalously bright and anomalously large.[35]

[35] In actual observations, the redshift will decrease the flux of photons. The effect of focussing by curvature would be clearly visible only if the observer moved so as to cancel the redshift and if there were no absorption of light between the emitter and the observer.

In order to make (13.53) a workable formula, we have to calculate the constant $C_{\mathcal{F}}$. Imagine then a sphere S with the centre at G and radius r_S.[36] Equation (13.53) shows that, for a given bundle of null geodesics, $(1+z)^2\mathcal{F}\delta S$ is constant along the bundle. On the surface of a sphere, δS is proportional to the solid angle $\delta\Omega_G$ subtended by δS, thus $\delta S_G = {r_S}^2 \delta\Omega_G$. Assuming that the source G radiates isotropically, and denoting by L its total luminosity (i.e. energy emitted per unit of time), we have on S $\mathcal{F}_S = L/(4\pi {r_S}^2)$. Also, $z = 0$ on S (because S is close to the source), so, taking (13.53) on S, we have

$$C_{\mathcal{F}} = \frac{L}{4\pi}\,\delta\Omega_G. \tag{13.54}$$

Now, following Etherington (1933) and Ellis (1971), we define the **source area distance** r_G (from the source to the observer) by

$$\delta S_o \overset{\text{def}}{=} {r_G}^2 \delta\Omega_G. \tag{13.55}$$

Finally, combining (13.53), (13.54) and (13.55) we obtain

$$\mathcal{F}_o = \frac{L}{4\pi}\,\frac{1}{(1+z_o)^2\,{r_G}^2}. \tag{13.56}$$

The quantity $\mathcal{F}_o$ can be measured (this is the flux of energy from a given source through a given surface sheet at the observer's position) and so can z_o. However, r_G cannot be measured because we cannot get close enough to the source to measure $\delta\Omega_G$. With L being essentially unknown, Eq. (13.56) cannot be observationally tested.

But we can define distance in another way. Imagine a surface sheet of area δS_G placed at the source G, and a bundle of light rays sent from this sheet in such a way that it converges to a single point at the observer's position O (Fig. 13.2). Assume that the sheet is orthogonal to the central ray of the bundle and that the central ray is emitted from G. Let the solid angle filled by this bundle as it reaches O be $\delta\Omega_o$. We can then define the **observer area distance** from G to O by

$$\delta S_G \overset{\text{def}}{=} {r_o}^2 \delta\Omega_o. \tag{13.57}$$

We can measure $\delta\Omega_o$, and we can *in principle* calculate δS_G (for example, assuming that the bundle subtends the whole source, whose geometric size we know). In this way, we can *in principle* measure r_o. It turns out that r_G is determined by r_o, as we show below.

13.5.2 The reciprocity theorem

Suppose that the galaxy G in Fig. 13.2 sends a light ray from its middle point that hits the observer O. It also sends a diverging bundle of rays that surrounds the central ray GO and, at G, fills the solid angle $\delta\Omega_G$. The bundle has the projected area δS_o at the observer's position. Another bundle of rays is sent from a surface sheet of area δS_G that

[36] We assume the radius of the sphere to be sufficiently small that the space inside it can be approximated by the flat tangent space to the manifold.

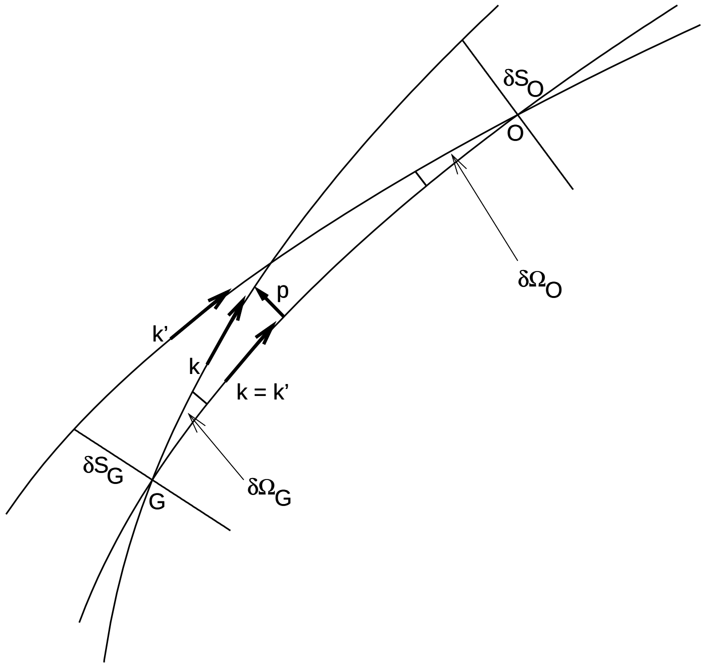

Figure 13.2: An illustration of the reciprocity theorem. Explanation is given in the text.

is orthogonal to GO at the central point G. The second bundle converges at O, and there it fills the solid angle $\delta\Omega_o$. (Figure 13.2 shows a cross-section through one half of each bundle.) Let the affine parameter and the field of null vectors tangent to the first bundle be v and k, respectively, and let the same quantities for the second bundle be v' and k'. Along the ray GO, $k = k'$. Let the field of geodesic deviation defined by the field k be p, and the deviation for the other bundle be p'. By construction, $p^\alpha = 0$ at G and $p'^\alpha = 0$ at O. Then the following is true:

Theorem 13.1 (the reciprocity theorem)

Assumption:

The bundle k completely surrounds O, i.e. there are rays intersecting δS_o in every direction from O.

Thesis:

$$r_G{}^2 = r_o{}^2 \left(1 + z_o\right)^2 . \tag{13.58}$$

Comment: The assumption would not be fulfilled if O were placed on a reflecting or absorbing surface, e.g. on a boundary between vacuum and opaque matter.

Proof:

On the central ray $v = v'$ and $k = k'$ (but not $p = p'$). Thus, the geodesic deviation equation (6.50) and the symmetries of the Riemann tensor imply that, along this ray,

$$p'^{\alpha}\ \frac{D^2 p_\alpha}{\mathrm{d}v^2} - p^\alpha\ \frac{D^2 p'_\alpha}{\mathrm{d}v^2} = 0. \tag{13.59}$$

Consequently

$$p'^{\alpha}\ \frac{D p_\alpha}{\mathrm{d}v} - p^\alpha\ \frac{D p'_\alpha}{\mathrm{d}v} = \text{constant} \tag{13.60}$$

along GO. Note that Eq. (6.48), with zero torsion, implies that $k^\beta p^\alpha{}_{;\beta} = p^\beta k^\alpha{}_{;\beta}$. Using this in (13.60) (where $Dp_\alpha/\mathrm{d}v = k^\beta p_{\alpha;\beta}$), we see that the connection terms cancel out. From now on we will thus replace the covariant derivatives by ordinary derivatives. Taking (13.60) at G (where $p^\alpha = 0$) and at O (where $p'^\alpha = 0$), we obtain

$$\left[p'^{\alpha}\ \frac{\mathrm{d}p_\alpha}{\mathrm{d}v}\right]_G = -\left[p^\alpha\ \frac{\mathrm{d}p'_\alpha}{\mathrm{d}v}\right]_o. \tag{13.61}$$

The geodesic deviation equation (6.50) determines only the component of δx^α transversal to k^μ, the component along k^μ obeys (6.50) identically and is not determined, also when k^μ is null. We may thus assume that $k_\alpha p^\alpha = k_\alpha p'^\alpha = 0$ all along the central ray GO. Of all these deviation vectors we now choose for further investigation those p^α that are orthogonal to the 4-velocity of the observer at O, and those p'^α that are orthogonal to the 4-velocity of the light source at G. The chosen vectors span 2-dimensional planes tangent to the spacetime at O and G. In the tangent plane at G there exist pairs that obey

$$\left[\frac{\mathrm{d}p_1{}^\alpha}{\mathrm{d}v}\ \frac{\mathrm{d}p_{2\alpha}}{\mathrm{d}v}\right]_G = 0. \tag{13.62}$$

They exist because of $p^\alpha|_G = 0$, so a given $(\mathrm{d}p^\alpha/\mathrm{d}v)_G$ selects one neighbouring geodesic from the k bundle, and the initial data $p^\alpha|_G = 0$, $(\mathrm{d}p^\alpha/\mathrm{d}v)_G$ uniquely determine k^α along the whole geodesic. By the assumption in the theorem, neighbouring geodesics exist in all directions around GO, so for any $(\mathrm{d}p_1^\alpha/\mathrm{d}v)_G$ there exists a $(\mathrm{d}p_2^\alpha/\mathrm{d}v)_G$ orthogonal to it that selects a neighbouring geodesic.

The vectors p_1 and p_2 whose initial data at G obey (13.62) will not in general be orthogonal at O, so let $\pi/2 + \psi$ be the angle between them at O. But the pairs obeying (13.62) can be rotated by any angle, by the assumption in the theorem. Suppose the pair is rotated by $\pi/2$. Then, arbitrarily near G, p_1 will become p_2, and p_2 will become $-p_1$. Since the solutions of the geodesic deviation equation with given initial data are unique, the same will happen at O: p_{1o} will become p_{2o}, and p_{2o} will become $-p_{1o}$. By elementary trigonometry, whatever the angle $\pi/2 + \psi$ was before the rotation, it will become $\pi/2 - \psi$ after the rotation. Consequently, by continuity, if the pair obeying (13.62) is rotated by all angles between 0 and $\pi/2$, then the corresponding pair at O will form all angles between $\pi/2 + \psi$ and $\pi/2 - \psi$, and one of those angles will be $\pi/2$. Choosing the pair for which it happens, we have

$$\left(p_1^\alpha p_{2\alpha}\right)\big|_o = 0. \tag{13.63}$$

From the field p'^α we now choose the vectors that at G obey

$$\left[p'^{\alpha}_1\ \frac{\mathrm{d}p_{2\alpha}}{\mathrm{d}v}\right]_G = 0 = \left[p'^{\alpha}_2\ \frac{\mathrm{d}p_{1\alpha}}{\mathrm{d}v}\right]_G. \tag{13.64}$$

Through (13.61) this implies that

$$\left[p_1^\alpha \frac{\mathrm{d}p'_{2\alpha}}{\mathrm{d}v}\right]_o = 0 = \left[p_2^\alpha \frac{\mathrm{d}p'_{1\alpha}}{\mathrm{d}v}\right]_o. \tag{13.65}$$

Since all the vectors lie in 2-planes at G and at O, the first equation of (13.64) and (13.62) imply that ${p'_1}^\alpha|_G$ is collinear with $\mathrm{d}p_1^\alpha/\mathrm{d}v|_G$; denote this by $p'_1|_G \propto \mathrm{d}p_1/\mathrm{d}v|_G$. Then the second of (13.64) says that

$$\left({p'_1}^\alpha p'_{2\alpha}\right)_G = 0. \tag{13.66}$$

By a similar consideration, (13.65) and (13.63) imply that

$$\left[\frac{\mathrm{d}{p'_1}^\alpha}{\mathrm{d}v} \frac{\mathrm{d}p'_{2\alpha}}{\mathrm{d}v}\right]_o = 0. \tag{13.67}$$

In consequence of (13.62), (13.63), (13.66) and (13.67), the norms of the vector products $\vec{p_1} \times \vec{p_2}$, $\vec{p'_1} \times \vec{p'_2}$, $\frac{\vec{\mathrm{d}p_1}}{\mathrm{d}v} \times \frac{\vec{\mathrm{d}p_2}}{\mathrm{d}v}$ and $\frac{\vec{\mathrm{d}p'_1}}{\mathrm{d}v} \times \frac{\vec{\mathrm{d}p'_2}}{\mathrm{d}v}$ reduce to products of the lengths of the vectors, denoted $|p_i|$, etc. Thus

$$\begin{aligned} \delta S_o &= [|p_1|\,|p_2|]_o\,, & \delta S_G &= [|p'_1|\,|p'_2|]_G\,, \\ \delta\Omega_G &= \left[\left|\frac{\mathrm{d}p_1}{\mathrm{d}\ell}\right| \cdot \left|\frac{\mathrm{d}p_2}{\mathrm{d}\ell}\right|\right]_G, & \delta\Omega_o &= \left[\left|\frac{\mathrm{d}p'_1}{\mathrm{d}\ell}\right| \cdot \left|\frac{\mathrm{d}p'_2}{\mathrm{d}\ell}\right|\right]_o, \end{aligned} \tag{13.68}$$

where ℓ is the radius at which the solid angles $\delta\Omega_G$ and $\delta\Omega_o$ subtend the surfaces $[|\mathrm{d}p_1|\,|\mathrm{d}p_2|]$ and $[|\mathrm{d}p'_1|\,|\mathrm{d}p'_2|]$, respectively. It is related to v by $\mathrm{d}\ell = -k^\rho u_\rho \mathrm{d}v$ (see the text after (13.28)). Thus we have from (13.68), (13.64) and (13.65):

$$\delta S_o \delta\Omega_o = \left[|p_1| \left|\frac{\mathrm{d}p'_1}{\mathrm{d}v}\right| |p_2| \left|\frac{\mathrm{d}p'_2}{\mathrm{d}v}\right|\right]_o \frac{1}{\left(k^\rho u_\rho\right)_o^{\,2}}. \tag{13.69}$$

We already know that $p'_1|_G \propto \mathrm{d}p_1/\mathrm{d}v|_G$. By the same method we conclude from (13.63) – (13.67) that $p'_2|_G \propto \mathrm{d}p_2/\mathrm{d}v|_G$, $p_1|_o \propto \mathrm{d}p'_1/\mathrm{d}v|_o$ and $p_2|_o \propto \mathrm{d}p'_2/\mathrm{d}v|_o$. Thus, the scalar products in all these pairs reduce to the products of their lengths. Consequently, (13.61) becomes

$$\left[|p'_1| \left|\frac{\mathrm{d}p_1}{\mathrm{d}v}\right|\right]_G = -\left[|p_1| \left|\frac{\mathrm{d}p'_1}{\mathrm{d}v}\right|\right]_o, \tag{13.70}$$

and the same is true for the subscript 2. Using this in (13.69) we obtain

$$\delta S_o \delta\Omega_o = \left[|p'_1| \left|\frac{\mathrm{d}p_1}{\mathrm{d}v}\right| |p'_2| \left|\frac{\mathrm{d}p_2}{\mathrm{d}v}\right|\right]_G \frac{1}{\left(k^\rho u_\rho\right)_o^{\,2}} = \frac{\left(k^\rho u_\rho\right)_G^{\,2}}{\left(k^\rho u_\rho\right)_o^{\,2}}\, \delta S_G \delta\Omega_G. \tag{13.71}$$

This is equivalent to (13.58) by (13.23), (13.55) and (13.57). □

Substituting (13.58) in (13.56) we obtain an expression for r_o:

$$r_o = \sqrt{\frac{L}{4\pi\mathcal{F}_o}}\, \frac{1}{(1+z)^2}. \tag{13.72}$$

This allows us to calculate r_o if we know L from theory and $\mathcal{F}_o$ from measurement. The r_o calculated in this way is called the **corrected luminosity distance**, while

$$D_L \overset{\text{def}}{=} \sqrt{\frac{L}{4\pi\mathcal{F}_o}} \tag{13.73}$$

is the ordinary (uncorrected) **luminosity distance**. It is clearly the distance from which the radiating body, if motionless in a Euclidean space, would produce an energy flux equal to the $\mathcal{F}_o$ measured by the observer. It is thus a fictitious quantity, meant to appeal to the imagination of the observer and unrelated to the geometry of the spacetime in question.

13.6 Exercises.

1. Prove that any vector orthogonal to two linearly independent null vectors is spacelike.

Hint: First prove that if for two null vectors k^α and ℓ^α we have $k^\alpha \ell_\alpha > 0$ then $v^\alpha \overset{\text{def}}{=} k^\alpha + \ell^\alpha$ is timelike (if $k^\alpha \ell_\alpha < 0$ then $u^\alpha \overset{\text{def}}{=} k^\alpha - \ell^\alpha$ is timelike). Then show that a vector w^α orthogonal to a timelike vector u^α must be spacelike. For example, decompose w^α in an orthogonal basis which includes u^α, then calculate $w^\alpha w_\alpha$ and use $w^\alpha u_\alpha = 0$.

2. Using the methods of Sec. 12.1 prove that the expansion scalar for a family of null geodesics in a flat space obeys

$$\theta = \frac{\mathrm{d}}{\mathrm{d}s} \ln S_n, \tag{13.74}$$

where s is the affine parameter on the null geodesics and S_n is the surface-area of the propagating light-front at the parameter value s.

3. Consider a ray proceeding from a spacetime point P_1 to P_2 and then from P_2 to P_3. Denote the redshifts acquired in the intervals $[P_1, P_2]$, $[P_2, P_3]$ and $[P_1, P_3]$ by z_{12}, z_{23} and z_{13}, respectively. Prove that (13.23) implies

$$1 + z_{13} = (1 + z_{12})\,(1 + z_{23})\,. \tag{13.75}$$