Is the Universe homogeneous?

BY ROY MAARTENS1,2,*

1Department of Physics, University of Western Cape, Cape Town 7535, South Africa
2Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK

The standard model of cosmology is based on the existence of homogeneous surfaces as the background arena for structure formation. Homogeneity underpins both general relativistic and modified gravity models and is central to the way in which we interpret observations of the cosmic microwave background (CMB) and the galaxy distribution. However, homogeneity cannot be directly observed in the galaxy distribution or CMB, even with perfect observations, since we observe on the past light cone and not on spatial surfaces. We can directly observe and test for isotropy, but to link this to homogeneity we need to assume the Copernican principle (CP). First, we discuss the link between isotropic observations on the past light cone and isotropic space–time geometry: what observations do we need to be isotropic in order to deduce space–time isotropy? Second, we discuss what we can say with the Copernican assumption. The most powerful result is based on the CMB: the vanishing of the dipole, quadrupole and octupole of the CMB is sufficient to impose homogeneity. Real observations lead to near-isotropy on large scales—does this lead to near-homogeneity? There are important partial results, and we discuss why this remains a difficult open question. Thus, we are currently unable to prove homogeneity of the Universe on large scales, even with the CP. However, we can use observations of the cosmic microwave background, galaxies and clusters to test homogeneity itself.

Keywords: cosmology; homogeneity; cosmic microwave background; galaxy distribution

1. Introduction

The standard model of the Universe—the Lambda cold dark matter (ΛCDM) ‘concordance’ model—is homogeneous, with structure formation described via perturbations. Given the assumption of homogeneity, and if general relativity (GR) correctly describes gravity, then the acceleration of the Universe is driven by dark energy. The homogeneous ΛCDM model is highly successful—a simple, predictive model that is compatible with all observations up to now. However, there is still no satisfactory description of the dark energy that is central to this model. This motivates the need to probe the foundations of the model. We can probe the assumption that GR holds on cosmological scales by investigating
modified gravity theories and by devising consistency tests of GR. This probe is only effective if we assume homogeneity. Alternatively, we can assume that GR holds and probe the assumption of homogeneity (see also [1,2]).

A common misconception is that ‘homogeneity is obvious from the cosmic microwave background (CMB) and the galaxy distribution’. In fact, we cannot directly observe or test homogeneity—since we observe down the past light cone, and not on spatial surfaces that intersect that light cone (figure 1). We only see the CMB on a 2-sphere at one redshift, and galaxy surveys give us the galaxy distribution on 2-spheres of constant redshift. There are interesting and important analyses of the observed galaxy distribution to probe statistical homogeneity (e.g. [3,4]), but these effectively assume a Friedmann–Lemaître–Robertson–Walker (FLRW) background geometry.

What we can directly test for is isotropy of observations. This then raises an important, but often overlooked, question: what observational quantities need to be isotropic in order to enforce isotropy of the geometry? This question is addressed in §2. To answer this question, we need a fully nonlinear analysis, since we cannot assume a priori any symmetric background space–time. For observations of the galaxy distribution, the answer is—we need isotropic angular distances, number counts, bulk velocities and lensing. Isotropy of the CMB by contrast does not in itself enforce space–time isotropy.

In order to link isotropy to homogeneity, we have to assume the Copernican principle (CP), i.e. that we are not at a special position in the Universe. The CP is not observationally based; it is an expression of the intrinsic limitation of observations from one space–time location. We consider in §3 what can be done with the CP. If we have isotropy along one worldline, based on the observed galaxy distribution, then the CP leads to homogeneity. A more powerful result is that homogeneity follows if all observers see isotropic angular distances up to third order in redshift. The strongest basis for homogeneity comes from the CMB. This is often considered to be obvious—but it is far from
Is the Universe homogeneous?

It is straightforward to show that homogeneity of the metric follows if all observers see isotropic CMB. The proof requires the general nonlinear Einstein–Liouville equations. Remarkably, it is not necessary to assume full CMB isotropy for each observer—it is enough that each observer sees isotropy in the CMB only up to the octupole. This is the most powerful observational basis for homogeneity currently known.

Of course, the real CMB is not exactly isotropic, but nearly isotropic. Does it follow from near-isotropy that the Universe is nearly homogeneous, i.e. perturbed FLRW? This has so far only been shown with further assumptions on the gradients and time derivatives of CMB multi-poles.

It is important to stress from the outset that there are two fundamental limitations:

— Isotropy and homogeneity of observables can only be meaningfully defined on large enough scales—and the nature of the transition scale is only poorly understood.
— Isotropy and homogeneity of observables can only be meaningfully defined in an average sense—and the problem of how to average in general relativity (and other metric theories that are intrinsically nonlinear) remains unsolved.

These unresolved issues are of crucial importance in cosmology, but they are not discussed here; instead, observations are treated as idealized.

2. What is the basis for isotropy?

Here we consider the situation when the CP is not adopted. What is the observational basis for showing that space–time is isotropic about the worldline of a single observer? Isotropy is directly observable and the best example is the CMB, which is isotropic about us to approximately $10^{-5}$ (after the dipole is interpreted as owing to our motion relative to the cosmic frame, and removed by a boost). Observations of the galaxy distribution do not have the same precision, but there is no evidence for anisotropy. First, we look at observations of matter and then of the CMB.

(a) Isotropic matter distribution on the past light cone of one observer

The dominant cosmological components—cold dark matter (CDM) and dark energy—have not been independently observed. Unlike baryonic matter, the dark components are up to now only manifest via their gravitational effect. The distribution of dark matter is mapped by weak-lensing surveys. But to relate the measured projected potential on the sky at each redshift to the dark matter, we require a specific model, such as a perturbed FLRW model. The dark matter 4-velocity is usually assumed to be aligned with that of baryonic matter—but this is also based on a perturbed FLRW model.

This unavoidably means that we must impose a model for these dark components—not merely their physical properties, but how they relate spatially to observed matter—in order to determine their distribution via cosmological observations. A starting point [1] is to assume that the CDM 4-velocity is the same.
as the baryonic 4-velocity, and that we know the primordial ratio of CDM density to baryonic density, as well as the bias factor that relates the concentrations of CDM and baryons in clustered matter

\[ \rho_c \text{ known from } \rho_b \quad \text{and} \quad u_0^a = u_c^a := u^a. \quad (2.1) \]

If there is a modified gravity theory that avoids the need for CDM, then we do not need (2.1)—but we are likely to need other assumptions on the extra degrees of freedom that mimic CDM.

If dark energy is in the form of \( \Lambda \), then we need to assume that its value is known from non-cosmological physics,

\[ \Lambda \text{ known independently of cosmological observations.} \quad (2.2) \]

For quintessence and other more complicated forms of dark energy, we would need to assume how the dark energy field is distributed in space–time—since we are not assuming an FLRW geometry \textit{a priori}. If there is a modified gravity theory that avoids the need for dark energy, then statement (2.2) is not needed, but assumptions will probably be necessary on the extra degrees of freedom that mimic dark energy.

Given the assumptions on the dark components, what can we say about space–time geometry if the matter distribution is isotropic on the past light cone of the observer? Which observables need to be isotropic in order to deduce isotropic geometry? It turns out that four independent observables on the light cone are exactly enough to impose isotropy of space–time. The original result for a baryonic Universe [5–7] may be updated to include CDM and \( \Lambda \) [1], incorporated via equations (2.1) and (2.2):

\[ \text{Isotropy of matter distribution on the light cone} \rightarrow \text{isotropy of space–time geometry} \]

If one fundamental observer co-moving with the matter measures isotropy of (i) angular diameter distances, (ii) number counts, (iii) bulk velocities, and (iv) lensing, in an expanding dust Universe with \( \Lambda \), then the space–time is isotropic about the observer’s worldline.

Note that isotropy of bulk velocities is equivalent to vanishing transverse velocities (proper motions) on the observer’s sky. Isotropy of lensing means that there is no distortion of images, only magnification.

The proof of this result requires a non-perturbative approach—there is no background to perturb around. Since the data are given on the past light cone of the observer, we need the full general metric, adapted to the past light cones of the observer worldline \( \mathcal{C} \). We define observational coordinates \( x^\mu = (w, y, \theta, \phi) \), where \( x^P = (\theta, \phi) \) are the celestial coordinates, \( w = \text{const.} \) are the past light cones on \( \mathcal{C} (y = 0) \), normalized so that \( w \) measures proper time along \( \mathcal{C} \), and \( y \) measures distance down the light rays \( (w, \theta, \phi) = \text{const.} \) (figure 2). A convenient choice for \( y \) is \( y = z \) (redshift) on the light cone of here and now, \( w = w_0 \), and then keep \( y \) co-moving with matter off the initial light cone, so that \( u_y = 0 \). Then the matter 4-velocity and the photon wavevector are

\[ u^\mu = (1 + z)(1, 0, V^P), \quad k_\mu = w_\mu, \quad 1 + z = u_\mu k^\mu, \quad (2.3) \]

where \( V^P = \frac{dx^P}{dw} \) are the transverse velocity components on the observer’s sky.
Is the Universe homogeneous?

The metric of an arbitrary space–time in observational coordinates is
\[ ds^2 = -A^2 dw^2 + 2B dw dy + 2C_P dx^P dw + D^2 (d\Omega^2 + L_{PQ} dx^P dx^Q) \] (2.4)
and
\[ A^2 = (1+z)^{-2} + 2C_P V^P + g_{PQ} V^P V^Q, \quad B = \frac{dv}{dy}, \] (2.5)
where \( v \) is a null affine parameter, \( D \) is the angular diameter distance, and \( L_{PQ} \) determines the lensing distortion of images via the shear of light rays,
\[ \hat{\sigma}_{PQ} = \frac{D^2 \partial L_{PQ}}{2B \partial y}. \] (2.6)

The number of galaxies in a solid angle \( d\Omega \) and a null distance increment \( dv = Bdy \) involves metric component \( B \),
\[ dN = S n (1+z) D^2 B d\Omega dy, \] (2.7)
where \( S \) is the selection function and \( n = \rho_m / m \) is the number density.

Before specializing to isotropic observations, we identify how the observations in general and in principle determine the geometry of the past light cone \( w = w_0 \) of here and now, where \( y = z \):

— Given the intrinsic properties and evolution of sources, observations in principle determine: (i) the angular diameter distance, \( D(w_0, z, x^P) \), and (ii) the lensing distortion of images, \( L_{PQ}(w_0, z, x^R) \). Thus, the metric components \( g_{PQ} \) on \( w = w_0 \) are determined.
— Given equation (2.1) and the selection function, the number counts, \( N(w_0, z, x^P) \), in principle determine \( B(w_0, z, x^P) \rho_m(w_0, z, x^P) \).

— In principle, observations over extended time scales determine the instantaneous transverse velocities, \( V^P(w_0, z, x^Q) \), of discrete sources. Transverse velocities of clusters are in principle determined by the polarization of scattered CMB photons [8].

It follows that, in principle and for idealized observations:

\[ \text{Idealized data} \Rightarrow \{u^\mu, B\rho_m, g_{PQ}\} \text{ on } w = w_0. \]  

But this is insufficient to determine the geometry of the past light cone, because we need \( g_{0P} = C_P \) and we cannot separate out \( B \) and \( \rho_m \). Without gravitational field equations, we are unable to fully determine the space–time and matter on the past light cone, even assuming perfect information from discrete-source observations. As a consequence, it is also impossible to test gravity theories directly:

\textit{Cosmological testing of gravity theories}

Even with perfect observations, we cannot determine the space–time geometry and matter on our past light cone without gravitational field equations. Thus observations cannot directly test GR on cosmological scales, or test any alternative theories of large-scale gravity, without making assumptions about the space–time geometry.

If GR holds, then (2.8) is exactly what is needed [5,6]:

\textit{Matter observations \( \rightarrow \) metric and matter on light cone}

Observational data (2.8) is exactly the information needed for Einstein’s equations to determine \( B, C_P \) on \( w = w_0 \), so that the metric and matter are fully determined on the light cone. Then the past Cauchy development of these data determines \( g_{\mu\nu}, u^\mu, \rho_m \) in the interior of the past light cone.

If the matter observations are isotropic, then we can prove isotropy of space–time and matter [5,6]:

\textit{Isotropy of light cone matter distribution for one observer \( \rightarrow \) LTB}

In an expanding dust region with \( \Lambda \), if one fundamental observer measures isotropic angular diameter distances, number counts, bulk velocities and lensing,

\[ \frac{\partial D}{\partial x^P} = \frac{\partial N}{\partial x^P} = V^P = L_{PQ} = 0, \]

then the space–time and matter are isotropic, i.e. the region is Lemaitre–Tolman–Bondi (LTB).

It is an open question how this result translates to the case of almost-isotropy of observations: does this lead to almost-isotropy of the space–time?
Is the Universe homogeneous?

(b) Isotropic cosmic microwave background for one observer

If the CMB is isotropic for one observer co-moving with the matter, then along the observer’s worldline $C$ we have (see equation (3.7))

$$f(x_C, p) = F(x_C, E), \quad F_{\mu_1 \ldots \mu_\ell} |_C = 0 = \dot{F}_{\mu_1 \ldots \mu_\ell} |_C \quad \text{for} \quad \ell \geq 1. \quad (2.10)$$

In other words, all covariant multi-poles of the distribution function beyond the monopole (see equation (A 32)), and their time derivatives, must vanish along $C$. By equation (A 33), the radiation momentum density (from the dipole) and anisotropic stress (from the quadrupole) vanish: $q^\mu |_C = 0 = \pi^{\mu \nu} |_C$. However, without the Copernican assumption, we are not able to deduce directly the vanishing of spatial derivatives $\tilde{V}^n F_{\mu_1 \ldots \mu_\ell} |_C$, and then we cannot show isotropy of the space–time geometry about $C$.

Isotropy of the CMB alone is not sufficient, since the matter could in principle be anisotropic, even if this is physically unnatural. In order to rule out artificial counter-examples, it would be sufficient to characterize the minimal matter isotropy that combines with CMB isotropy to give geometric isotropy.

If we adopt the CP, then the CMB alone leads to a powerful result, as discussed in §3.

3. What is the basis for spatial homogeneity?

Homogeneity cannot be directly observed—we are effectively unable to move away in cosmic time or distance from here and now and hence cannot probe spatial variations on constant time slices; effectively, our observations only access the past light cone of here and now. Direct observation thus cannot distinguish between an evolving homogeneous distribution of matter and inhomogeneity with a different time evolution—since the past light cone only accesses a 2-sphere in each constant-time slice (figure 1). Thus, we are forced to adopt the Copernican assumption. We first consider matter observations, then an exactly isotropic CMB, and finally the case of an almost-isotropic CMB.

(a) Isotropic matter observations for all observers

If all observers see isotropy of the four matter observables, then we have geometric isotropy along all worldlines and thus homogeneity follows.

Isotropy of light cone matter distribution for all observers $\rightarrow$ FLRW

In an expanding dust region with $A$, if all fundamental observers measure isotropic angular diameter distances, number counts, bulk velocities and lensing, then the space–time is FLRW.

This is an improved form of the cosmological principle—based on isotropy of specific observables and not on assumed geometric isotropy.

In fact there is a much stronger statement than this, based only on one observable, the angular diameter distance, and only for small redshifts [9]:

Phil. Trans. R. Soc. A (2011)
Isotropic distances up to $O(z^3)$ for all observers $\rightarrow$ FLRW

If all fundamental observers in an expanding space–time region measure isotropic angular diameter distances up to third order in a redshift series expansion, then the space–time is FLRW in that region.

A covariant proof of this [1,10] is based on a series expansion in a general space–time, using the method of Kristian & Sachs [11]. The redshift may be expanded in terms of the angular diameter distance,

$$z = \left[ k^\mu k^\nu \nabla_\mu u_\nu \right]_O D + \frac{1}{2} \left[ k^\mu k^\nu k^\alpha \nabla_\alpha u_\nu \right]_O D^2$$

$$+ \frac{1}{6} \left[ k^\mu k^\nu k^\alpha k^\beta \nabla_\alpha \nabla_\mu u_\beta + \frac{1}{2} k^\mu k^\nu k^\alpha k^\beta R_{\alpha\beta} \nabla_\mu u_\nu \right] D^3 + O(D^4). \tag{3.1}$$

Here all terms are evaluated at the observer $O$ in the unit direction $e^\mu$ of the light ray, where (see equation (2.3))

$$k^\mu = -(1 + z)(u^\mu + e^\mu), \quad u_\mu e^\mu = 0 \quad \text{and} \quad e_\mu e^\mu = 1. \tag{3.2}$$

The covariant derivative of the 4-velocity is decomposed as in (A 5). The $O(D)$ coefficient is the observed Hubble rate (generalizing the FLRW quantity),

$$H_0^{\text{obs}} = \left[ k^\mu k^\nu \nabla_\mu u_\nu \right]_O = \left[ \frac{1}{3} \Theta + A_\mu e^\mu + \sigma_{\mu\nu} e^\mu e^\nu \right]_O. \tag{3.3}$$

Isotropy at lowest order, for all observers $O$, thus enforces vanishing acceleration and shear

$$A_\mu = 0 = \sigma_{\mu\nu}. \tag{3.4}$$

Using equation (3.4), the next coefficient becomes

$$\left[ k^\mu k^\nu k^\alpha \nabla_\mu \nabla_\nu u_\alpha \right]_O = \frac{1}{6} \left[ (2\Theta^2 + \rho_m - 2A) - 2\tilde{\nabla}_\mu \Theta e^\mu + (E_{\mu\nu} + \omega_{\mu} \omega_{\nu}) e^\mu e^\nu \right]_O, \tag{3.5}$$

where $\tilde{\nabla}_\mu$ is the covariant spatial derivative (A 3), $E_{\mu\nu}$ is the electric part of the Weyl tensor (A 8), and the angled brackets denote the spatial tracefree part (A 2). Isotropy at $O(D^2)$, therefore, imposes homogeneity of the expansion rate, $\tilde{\nabla}_\mu \Theta = 0$, and the condition $E_{\mu\nu} = -\omega_{\mu} \omega_{\nu}$. However, this condition is identically satisfied by virtue of equation (3.4) and the shear propagation equation (A 14). The constraint equations (A 18) and (A 19) then show that $\text{curl} \omega_\mu = 0$ and $H_{\mu\nu} = \tilde{\nabla}_\mu \omega_{\nu}$, where $H_{\mu\nu}$ is the magnetic Weyl tensor. The $O(D^3)$ coefficient imposes further constraints, and we find that $0 = \tilde{\nabla}_\mu \rho_m = \omega_\mu = E_{\mu\nu} = H_{\mu\nu}$. Putting everything together, we have a covariant characterization of FLRW.

(b) Exactly isotropic cosmic microwave background for all observers

It is commonly assumed that isotropy of the CMB for all observers leads obviously to FLRW, without the need for any proof. In fact, the proof is far from obvious, and requires a nonlinear analysis of the general Einstein–Liouville equations. The starting point is a pioneering mathematical result by
Ehlers–Geren–Sachs [12]. They assumed that the only source of the gravitational field was the radiation, i.e. they neglected matter and $A$. This can be generalized to include self-gravitating matter and dark energy [1] (extending [13–18]):

\[ CMB \text{ isotropy for all observers } \rightarrow FLRW \]

In a region, if
— collisionless radiation is exactly isotropic,
— the radiation 4-velocity is geodesic and expanding, and
— there is dust matter and dark energy in the form of $A$, quintessence or a perfect fluid,

then the metric is FLRW in that region.

The fundamental four-velocity $u^\mu$ is the radiation 4-velocity, which has zero 4-acceleration and positive expansion,

\[ A_\mu = 0 \quad \text{and} \quad \Theta > 0. \]  

(3.6)

Isotropy of the radiation distribution about $u^\mu$ means that photon peculiar velocities are isotropic for co-moving observers; thus in momentum space, the photon distribution depends only on components of the 4-momentum $p^\mu$ along $u^\mu$, i.e. on the photon energy $E = -u_\mu p^\mu$,

\[ f(x, p) = F(x, E) \quad \text{and} \quad F_{\mu_1 \cdots \mu_\ell} = 0 \quad \text{for } \ell \geq 1. \]  

(3.7)

In other words, all covariant multi-poles of the distribution function beyond the monopole, defined in equation (A 32), must vanish. In particular, equation (A 33) shows that

\[ q_{\mu}^{\nu} = 0 = \pi_{\mu \nu}^{\text{r}}. \]  

(3.8)

Equation (3.7) also implies that the radiation brightness octupole $\Pi_{\mu\nu\alpha}$ and hexadecapole $\Pi_{\mu\nu\alpha\beta}$ are zero. These are source terms in the anisotropic stress evolution equation, which is the $\ell = 2$ case of equation (A 37). The general nonlinear form of the $\pi_{\mu \nu}^{\text{r}}$ evolution equation is [13,19]

\[
\begin{align*}
\dot{\pi}_{\mu \nu}^{\text{r}} & + \frac{4}{3} \Theta \pi_{\mu \nu}^{\text{r}} + \frac{8}{15} \rho_{\nu} \pi_{\mu}^{\nu r} + \frac{2}{5} \nabla_{\mu} (\dot{q}_{r}^{\nu p}) + 2 A_{\mu} (\dot{q}_{r}^{\nu p}) - 2 \omega^{p}_{\sigma \alpha \beta} (\mu \pi_{\nu}^{p})^{\beta} \\
& + \frac{2}{7} \sigma_{\alpha}^{(\mu \pi_{\nu}^{p})^{\alpha}} + \frac{8}{35} \nabla_{\alpha} \Pi_{\mu \nu}^{\text{r}} - \frac{32\pi}{315} \sigma_{\alpha \beta} \Pi_{\mu \nu}^{\text{r}} = 0.
\end{align*}
\]  

(3.9)

Isotropy removes all terms on the left except the third, and thus enforces a shear-free expansion of the fundamental congruence,

\[ \sigma_{\mu \nu} = 0. \]  

(3.10)

We can also show that $u^\mu$ is irrotational as follows. Together with equation (3.6), momentum conservation for radiation, i.e. equation (A 30) with $I = r$, reduces to

\[ \nabla_{\mu} \rho_{r} = 0. \]  

(3.11)

Thus, the radiation density is homogeneous relative to fundamental observers. Now we invoke the exact nonlinear identity for the covariant curl of the
gradient (A 6),
\[ \text{curl} \tilde{\nabla}_\mu \rho_t = -2 \dot{\rho}_t \omega_\mu \Rightarrow \Theta \rho_t \omega_\mu = 0, \] (3.12)
where we have used the energy conservation equation (A 29) for radiation. By assumption \( \Theta > 0 \), and hence we deduce that the vorticity must vanish,
\[ \omega_\mu = 0. \] (3.13)

Then we see from the curl shear constraint equation (A 19) that the magnetic Weyl tensor must vanish,
\[ H_{\mu \nu} = 0. \] (3.14)

Furthermore, equation (3.11) actually tells us that the expansion must also be homogeneous. From the radiation energy conservation equation (A 29), and using equation (3.8), we have \( \Theta = -3 \dot{\rho}_t / 4 \rho_t \). On taking a covariant spatial gradient and using the commutation relation (A 7), we find
\[ \tilde{\nabla}_\mu \Theta = 0. \] (3.15)
Then the shear divergence constraint, equation (A 18), enforces the vanishing of the total momentum density in the fundamental frame,
\[ q_\mu \equiv \sum_I q_\mu^I = 0 \Rightarrow \sum_I \gamma^2_I (\rho^*_I + p^*_I) v_\mu^I = 0. \] (3.16)
The second equality follows from equation (A 26), using the fact that the baryons, CDM and dark energy (in the form of quintessence or a perfect fluid) have vanishing momentum density and anisotropic stress in their own frames, i.e.
\[ q^*_\mu = 0 = p^*_\mu, \] (3.17)
where the asterisk denotes the intrinsic quantity (see appendix A). If we include other species, such as neutrinos, then the same assumption applies to them. Except in artificial situations, it follows from equation (3.16) that
\[ v_\mu^I = 0, \] (3.18)
i.e. the bulk peculiar velocities of matter and dark energy (and any other self-gravitating species satisfying (3.17)) are forced to vanish—all species must be co-moving with the radiation.

The co-moving condition (3.18) then imposes the vanishing of the total anisotropic stress in the fundamental frame,
\[ \pi^{\mu \nu} \equiv \sum_I \pi^{\mu \nu}_I = \sum_I \gamma^2_I (\rho^*_I + p^*_I) v^{(\mu}_I v^{\nu)}_I = 0, \] (3.19)
where we used equations (A 27), (3.17) and (3.18). Then the shear evolution equation (A 14) leads to a vanishing electric Weyl tensor
\[ E_{\mu \nu} = 0. \] (3.20)

Equations (3.6) and (3.19) now lead, via the total momentum conservation equation (A 12) and the \( E \)-divergence constraint (A 20), to homogeneous total
density and pressure,

$$\tilde{\nabla}_\mu \rho = 0 = \tilde{\nabla}_\mu p.$$  \hfill (3.21)

Equations (3.6), (3.10), (3.14)–(3.16), (3.19) and (3.21) constitute a covariant characterization of an FLRW space–time. This establishes the Ehlers–Geren–Sachs result, generalized from the original to include self-gravitating matter and dark energy, and presented in a fully covariant form. It is straightforward to include other species such as neutrinos. The critical assumption needed for all species is the vanishing of the intrinsic momentum density and anisotropic stress, i.e. equation (3.17). Equivalently, the energy–momentum tensor for the $I$-component should have perfect fluid form in the $I$-frame. The isotropy of the radiation and the geodesic nature of its 4-velocity then enforce the vanishing of (bulk) peculiar velocities $v^I_I$. We emphasize that one does not need to assume that the matter or other species are co-moving with the radiation—it follows from the assumptions on the radiation.

In fact this result can be dramatically strengthened on the basis of a theorem by Ellis–Treciokas–Matravers [20]: we do not need vanishing of all multi-poles, but only the dipole, quadrupole and octupole! The key step is to show that the shear vanishes, without having zero hexadecapole—the quadrupole evolution equation (3.9) no longer automatically gives $\sigma_{\mu\nu} = 0$, and we need to find another way to show this. The trick is to return to the Liouville multi-pole equation (A 36). The $\ell = 2$ multi-pole of this equation, with $F_\mu = F_\mu = F_\mu$, gives

$$\frac{12}{63} \frac{\partial}{\partial E} (E^5 \sigma_{\alpha\beta} F_{\mu\nu\alpha}) + E^5 \frac{\partial F}{\partial E} \sigma_{\alpha\beta} = 0.$$  \hfill (3.22)

We integrate over $E$ from 0 to $\infty$, and use the convergence property $E^5 F_{\mu\nu\alpha} \to 0$ as $E \to \infty$. This gives

$$\sigma_{\alpha\beta} \int_0^\infty E^5 \frac{\partial F}{\partial E} dE = 0.$$  \hfill (3.23)

Integrating by parts, the integral reduces to $-\frac{5}{4} \int_0^\infty E^4 F dE$. Since $F > 0$, the integral is strictly negative, and thus we arrive at vanishing shear, $\sigma_{\mu\nu} = 0$. Then the proof above proceeds as before, incorporating matter and dark energy to extend the original result [1]:

**CMB partial isotropy for all observers $\to$ FLRW**

In a region, if

— collisionless radiation has vanishing dipole, quadrupole and octupole,

$$F_\mu = F_\mu = F_\mu = 0,$$  \hfill (3.24)

— the radiation four-velocity is geodesic and expanding, and

— there is dust matter and dark energy in the form of $A$, quintessence or a perfect fluid,

then the metric is FLRW in that region.

This is the most powerful observational basis that we have for background homogeneity and thus an FLRW background model.
(c) The real universe: almost-isotropic cosmic microwave background

In practice, we can only observe approximate isotropy (figure 3). Is the previous result stable—i.e. does almost-isotropy of the CMB lead to an almost-FLRW Universe? This would be the realistic basis for a spatially homogeneous Universe (assuming the CP). It was shown to be the case, but subject to further assumptions, by Stoeger et al. [13] and Maartens et al. [21]:

A more realistic basis for homogeneity

In a region of an expanding Universe with dust and cosmological constant, if all observers co-moving with the matter measure an almost isotropic distribution of collisionless radiation, and if some of the time and spatial derivatives of the covariant multi-poles are also small, then the region is almost FLRW.

We emphasize that the perturbative assumptions are purely on the photon distribution, not the matter or the metric—and one has to prove that the matter and metric are then perturbatively close to FLRW. Once again, a non-perturbative analysis is essential, since we are trying to prove an almost-FLRW space–time, and we cannot assume it a priori.

Almost-isotropy of the photon distribution means that

$$F_{\mu_1 \ldots \mu_\ell}(x, E) = \mathcal{O}(\epsilon), \quad \ell \geq 1,$$

where $\epsilon$ is a (dimensionless) smallness parameter. The brightness multi-poles $P_{M_\ell}$ have dimensions of energy density and we, therefore, normalize them to the monopole $P = \rho_0/4\pi$, so that $P_{M_\ell}/P = \mathcal{O}(\epsilon)$.

The task is to show that the relevant kinematical, dynamical and curvature quantities, suitably non-dimensionalized, are $\mathcal{O}(\epsilon)$. For example, the dimensionful kinematical quantities may be normalized by the expansion, $\sigma_{\mu\nu}/\Theta, \sigma_{\mu}/\Theta$.
Is the Universe homogeneous?

The proof then follows the same pattern as the proof above of the exact result—except that, at each stage, we need to show that quantities are $O(\epsilon)$ rather than equal to zero.

However, in order to show this, we need smallness not just of the multi-poles but also of some of their derivatives. Smallness of the multi-poles does not directly imply smallness of their derivatives, and we have to assume this [18,22]. It remains a difficult open problem whether these additional assumptions may be removed. If all observers measure small multi-poles, then it may be possible to use almost-isotropy of the matter distribution to show that the time and space derivatives on cosmologically significant scales must also be small.

A number of experiments have been proposed to test the CP by looking for violations of isotropy at events down our past light cone, as discussed in the next section. The almost-isotropic CMB result then gives a framework for probing inhomogeneities via such observations. Indeed, these tests may provide a way of constraining spatial gradients of the low-$\ell$ multi-poles.

In addition, it may be possible to strengthen the almost-isotropic result above by proving that it is sufficient for only the first three multi-poles and their derivatives to be small. This would represent a more realistic foundation for almost-homogeneity.

4. Testing homogeneity

Although we cannot directly probe homogeneity by observations, we can test for violations of homogeneity. If we find no violation, then the indirect evidence for homogeneity is strengthened. However, if even one significant violation is discovered, then homogeneity will have been disproved. There are two broad classes of observational tests of homogeneity, one based on consistency relations that hold in FLRW and the other based on using galaxy clusters as probes of the anisotropy seen in the CMB from positions at cosmological distances down our past light cone.

(a) Consistency of distances and expansion rate

The effective standard candles provided by supernovae observations lay the basis for a consistency test of homogeneity. There are two geometric effects on distance measurements: the curvature bends null geodesics and the expansion changes radial distances. These are coupled in FLRW models via

$$D_L(z) = \frac{(1 + z)}{H_0\sqrt{-\Omega_{K0}}} \sin \left( \sqrt{-\Omega_{K0}} \int_0^z \frac{dz'}{H(z')/H_0} \right),$$

and then one can combine the Hubble rate and distance data to find the curvature today,

$$\Omega_{K0} = \frac{H^2(z)[(1 + z)D_L'(z) - D_L(z)]^2 - (1 + z)^4}{H_0^2(1 + z)^2D_L^2(z)}.$$

This relation is independent of all other cosmological parameters, including dark energy—and it is also independent of the theory of gravity. It can be used at a single redshift to determine $\Omega_{K0}$. Furthermore, it is the basis for a homogeneity
test, proposed by Clarkson et al. [23]: since $\Omega_{K0}$ is independent of $z$, we can differentiate to get the consistency relation,

$$K(z) := (1+z)^4 + H^2(z) \left[ (1+z)^2 \left\{ D_L(z) D''_L(z) - D''_L(z) \right\} + D''_L(z) \right]$$

$$+ (1+z) H(z) H'(z) D_L(z) [(1+z) D'_L(z) - D_L(z)]$$

$$= 0 \quad \text{for FLRW geometry. (4.3)}$$

$K = 0$ for any FLRW geometry, independent of curvature, dark energy, matter content and theory of gravity. In realistic models, we should expect $|K(z)| \sim 10^{-5}$, reflecting perturbations from large-scale structure formation. Significantly larger values indicate a breakdown of homogeneity,

$$K(z) \text{ significantly different from } 0 \Rightarrow \text{non-FLRW universe.}$$

Carrying out this test is no more difficult than carrying out dark energy measurements of $w(z)$ from supernovae type 1a data, which require $H'(z)$ from distance measurements or the second derivative $D''_L(z)$.

This is the simplest test of homogeneity, and its implementation should be regarded as a high priority. Another test involves the time drift of the cosmological redshift [24], but will only be feasible on a much longer timescale.

Finally the baryon acoustic oscillation (BAO) feature itself provides in principle a homogeneity test. Future large-volume surveys will allow the detection of the BAO scale in both radial and transverse directions. The physical lengths in radial and transverse directions of a feature with redshift extent $\Delta z$ and subtending an angle $\Delta \theta$ are

$$L_\parallel = \frac{\Delta z}{(1+z) H_\parallel(z)}, \quad L_\perp = D(z) \Delta \theta,$$

where $D$ is the angular diameter distance and $H_\parallel$ is the expansion rate in the radial direction (figure 4). The two scales are equal for the BAO feature in an FLRW background, where expansion is isotropic at all points. Any significant disagreement between the radial and transverse BAO scales would signal a breakdown of remote isotropy and thus of homogeneity:

$$\frac{L_\parallel}{L_\perp} - 1 \text{ significantly different from } 0 \Rightarrow \text{non-FLRW universe.}$$

(b) Sunyaev–Zeldovich effect: temperature of scattered cosmic microwave background photons

Galaxy clusters with their hot ionized intra-cluster gas act via scattering of CMB photons like giant mirrors that carry information about the last scattering surface seen by the cluster (figure 5). In other words, clusters allow us in principle to indirectly probe inside our past light cone.

CMB photons are scattered into our line of sight, thus inducing spectral distortions in the CMB temperature that we observe [25,26]. The thermal Sunyaev–Zeldovich (SZ) effect from the thermal motion of electrons reflects the monopole seen by the cluster. If the blackbody temperatures of CMB photons arriving at the cluster from points inside our light cone are significantly different from the blackbody temperature that we directly observe, then there will be
Is the Universe homogeneous?

Figure 4. Radial and transverse BAO scales. (Online version in colour.)

Figure 5. CMB photons from inside our past light cone are scattered into our line of sight by the ionized gas in galaxy clusters, thus carrying information about the level of isotropy seen by the clusters. (Online version in colour.)

Phil. Trans. R. Soc. A (2011)
a significant distortion. Such a signal would indicate that the cluster sees a significantly anisotropic CMB, hence violating the CP and homogeneity [27]. (This test has been applied to a class of LTB models by Caldwell & Stebbins [28].)

The bulk radial motion of the cluster gas induces a kinetic SZ signal that reflects the CMB dipole seen by the cluster. If this is large, then there would be a violation of the CP and homogeneity. This has been applied to classes of LTB models by Garcia-Bellido & Haugbolle [29] and Zhang & Stebbins [30], but it is in fact more generally applicable as a test of homogeneity.

In summary,

Large (non-perturbative) thermal or kinetic SZ temperature effect
⇒ non-FLRW universe.

(c) Sunyaev–Zeldovich effect: polarization of scattered cosmic microwave background photons

Analogous to the SZ effect on CMB temperature, there is an SZ effect on CMB polarization [8]. The cluster bulk transverse velocity, and the CMB monopole, quadrupole and octupole, as seen by the cluster, induce modifications in CMB polarization via scattering off the cluster gas. These effects have been computed in FLRW models [31–33]. The SZ polarization signals potentially contain more information than the SZ temperature signals, and this has been proposed for the standard FLRW model as a way to map the CMB quadrupole seen at remote locations, thus lessening the cosmic variance [33–36].

From a more general viewpoint, the SZ effect on polarization is in principle a powerful probe of the CP and homogeneity. In principle, if CMB polarization data indicate large transverse cluster velocities, or large modifications to the primordial polarization signal, this could signal a violation of the CP and homogeneity:

Large (non-perturbative) SZ polarization effects ⇒ non-FLRW universe.

5. Conclusions

It is important to re-examine the basic assumptions of the standard concordance model, especially in view of the problems raised by the fine-tuned and unnatural nature of dark energy. Here, we have re-examined the central assumption of homogeneity. We argued that homogeneity cannot be established or directly confirmed by observations, given the inherent limitations of light cone-based data. By contrast, isotropy can be directly probed by observations. We constructed the observational coordinates necessary to answer the question of which observables on the past light cone are needed to prove isotropy. It turns out that four independent matter observables—angular diameter distances, number counts, lensing distortion and transverse velocities—are exactly sufficient to determine the space–time geometry. Isotropy of these four observables along one worldline imposes isotropy of the space–time geometry. Surprisingly, isotropy of the CMB along one worldline does not in itself lead to isotropic geometry.

To establish homogeneity, we are forced to adopt the CP. Isotropy of the four matter observables for all observers leads to homogeneity—giving a more
Is the Universe homogeneous?

observational version of the cosmological principle. A surprising and more powerful result is that isotropy of only the angular diameter distances for all observers, and only to $O(z^3)$, enforces homogeneity.

The strongest observational basis for homogeneity comes from the CP combined with the high isotropy of the CMB. The main additional assumption needed is that the CMB rest frame is geodesic. We outlined a covariant nonlinear proof and generalization of the original Ehlers–Geren–Sachs result. And we highlighted the remarkable Ellis–Treciokas–Matravers result, i.e. that the vanishing of the dipole, quadrupole and octupole is sufficient to enforce homogeneity. The realistic case, with almost-isotropy of the CMB, does not lead to almost-homogeneity without additional assumptions on some of the derivatives of the multi-poles. It remains an open problem whether these assumptions can be avoided, possibly using further information from almost-isotropy of matter observations.

Although we cannot directly observe homogeneity, we can test homogeneity, using observations that carry information from inside our past light cone. We described how this can be achieved via consistency relations between distances and expansion rates, using supernova and BAO data, and via SZ effects from galaxy clusters on the CMB temperature and polarization. Up to now, none of these tests has indicated a breakdown of the CP and thus a violation of homogeneity. But the further advance of high-precision data will provide new opportunities to apply and extend these critical tests.

In summary, the standard homogeneous model of cosmology is successful, predictive and up to now robust against all the observational data, and against current tests of homogeneity. The problem of a satisfactory explanation for dark energy may be resolved by advances in particle physics and quantum gravity. It is certainly reasonable to assume that the $\Lambda$CDM model is a good description of the Universe. Nevertheless, it is also necessary to continue probing the foundations of the model—not only the assumption of homogeneity, but also other critical questions, such as the problem of averaging and the need for a detailed understanding of light propagation in a lumpy Universe.

I thank Chris Clarkson, Catherine Cress, Ruth Durrer, George Ellis, Alan Heavens, Raul Jimenez and Obinna Umeh for discussions. I am supported by a South African SKA Research Chair, by the UK Science & Technology Facilities Council, and by a Royal Society (UK)/NRF (South Africa) exchange grant between the Universities of Portsmouth and Western Cape.

Appendix A. Nonlinear field and Boltzmann equations

For convenience, we collect the key equations in the 1 + 3 covariant Lagrangian formalism (see [1] and references therein). This formalism provides a physically transparent formulation of the field equations and the Boltzmann equation in full nonlinear generality. It is based on a decomposition relative to a chosen 4-velocity field $u^\alpha$. The fundamental tensors are

$$h_{\mu\nu} = g_{\mu\nu} + u_\mu u_\nu \quad \text{and} \quad \epsilon_{\mu\nu\alpha} = \eta_{\mu\nu\alpha\beta} u^\beta,$$

where $h_{\mu\nu}$ projects into the instantaneous rest space of co-moving observers, and $\epsilon_{\mu\nu\alpha}$ is the projection of the space–time alternating tensor $\eta_{\mu\nu\alpha\beta} =$
The projected symmetric tracefree (PSTF) parts of vectors and rank-2 tensors are

\[ V_{(\mu)} = h_\mu \nu V_\nu \quad \text{and} \quad S_{(\mu\nu)} = \left\{ h_{(\alpha} \h_\nu \beta) - \frac{1}{3} h^{\alpha\beta} h_{\mu\nu} \right\} S_{\alpha\beta}. \]  

The skew part of a projected rank-2 tensor is spatially dual to the projected vector, \( S_{\mu} = \frac{1}{2} \delta_{\mu}^{\alpha} \delta^{\nu} \h_\nu \), and then any projected rank-2 tensor has the decomposition

\[ S_{\mu\nu} = (1/3) S h_{\mu\nu} + \delta_{\mu\nu} S_{\alpha\beta} + S_{(\mu\nu)}, \]  

where \( S = S_{\alpha\beta} \).

The covariant derivative \( \nabla_{\mu} \) defines 1+3 covariant time and spatial derivatives,

\[ {\dot{J}}_{\mu} \cdots \cdots = u_a \nu a J_{\mu} \cdots \cdots \quad \text{and} \quad \tilde{\nabla}_a {\dot{J}}_{\mu} \cdots \cdots = h_{a}^{\alpha} h_{\mu}^{\beta} \cdots h_{\nu}^{\tau} \nu \delta J_{\kappa} \cdots \tau. \]  

The projected derivative \( \tilde{\nabla}_\mu \) defines a covariant PSTF divergence, \( \tilde{\nabla}_\mu V_\mu \), \( \tilde{\nabla}_\nu S_{\mu\nu} \), and a covariant PSTF curl,

\[ \text{curl} \; V_\mu = \epsilon_{\mu\nu\rho} \tilde{\nabla}^\rho V_\nu \quad \text{and} \quad \text{curl} \; S_{\mu\nu} = \epsilon_{\alpha\beta(\mu} \tilde{\nabla}_{\nu)} S_{\rho)\beta}. \]  

The relative motion of co-moving observers is encoded in the PSTF kinematical quantities: the volume expansion rate, 4-acceleration, vorticity and shear, given respectively by

\[ \Theta = \tilde{\nabla}_\mu u_\mu, \quad A_\mu = \dot{u}_\mu, \quad \omega_\mu = \text{curl} \; u_\mu \quad \text{and} \quad \sigma_{\mu\nu} = \tilde{\nabla}_{(\mu} u_{\nu)} \]  

\[ = \frac{1}{3} \Theta h_{\mu\nu} + \epsilon_{\mu\nu\rho} \omega^\rho + \sigma_{\mu\nu} - A_\mu u_\nu. \]  

Key nonlinear identities are

\[ \text{curl} \; \tilde{\nabla}_\mu \psi := \epsilon_{\mu\rho} \tilde{\nabla}^\rho \tilde{\nabla}^\alpha \psi = -2 \psi \omega_\mu, \]  

\[ h_{\mu}^{\nu} (\tilde{\nabla}_\nu \psi) - \tilde{\nabla}_\mu \dot{\psi} = \dot{\psi} A_\mu - \left( \frac{1}{3} \Theta h_{\mu\nu} + \sigma_{\mu\nu} + \epsilon_{\mu\nu\rho} \omega^\rho \right) \tilde{\nabla}_\nu \psi. \]  

The PSTF dynamical quantities describe the sources of the gravitational field: the (total) energy density \( \rho = T_{\mu\nu} u_\mu u_\nu \), isotropic pressure \( p = (1/3) h_{\mu\nu} T^{\mu\nu} \), momentum density \( q_\mu = -T_{(\mu\nu)} u_\nu \) and anisotropic stress \( \pi_{\mu\nu} = T_{(\mu\nu)} \), where \( T_{\mu\nu} \) is the total energy–momentum tensor. The Weyl tensor splits into the PSTF gravito-electric and gravito-magnetic fields

\[ E_{\mu\nu} = C_{\mu\rho\beta} u_\rho u_\beta \quad \text{and} \quad H_{\mu\nu} = \frac{1}{2} \epsilon_{\mu\alpha\beta} C^{\alpha\beta}_{\nu\rho} u_\rho. \]  

The Ricci and Bianchi identities,

\[ \nabla_{[\mu} \nabla_{\nu]} u_\alpha = R_{\mu\nu\alpha\beta} u_\beta \quad \text{and} \quad \nabla^\beta C_{\mu\nu\alpha\beta} = -\nabla_{[\mu} \left\{ R_{\nu]a} - \frac{1}{6} R g_{\nu a} \right\}, \]  

produce the fundamental evolution and constraint equations governing the covariant quantities. Einstein’s equations are incorporated via the algebraic
replacement $R^{\mu \nu} = T^{\mu \nu} - \frac{1}{2} T_{\alpha}^{\alpha} g^{\mu \nu}$. The resulting equations, in fully nonlinear form and for a general source of the gravitational field, are:

**Evolution**

\[
\dot{\rho} + (\rho + p) \Theta + \tilde{\nabla}^{\mu} q_\mu = -2 A^{\mu} q_\mu - \sigma^{\mu \nu} \pi_{\mu \nu}, \quad (A\ 10)
\]

\[
\dot{\Theta} + \frac{1}{3} \Theta^2 + \frac{1}{2} (\rho + 3 p) - \tilde{\nabla}^{\mu} A_\mu = -\sigma_{\mu \nu} \sigma^{\mu \nu} + 2 \omega_\mu \omega^\mu + A_\mu A^\mu, \quad (A\ 11)
\]

\[
\dot{q}_{(\mu)} + \frac{4}{3} \Theta q_\mu + (\rho + p) A_\mu + \tilde{\nabla}_\mu p + \tilde{\nabla}^{\nu} \pi_{\mu \nu} = -\sigma_{\mu \nu} q^\nu + \epsilon_{\mu \nu \alpha} \omega^\nu q^\alpha - A^\nu \pi_{\mu \nu}, \quad (A\ 12)
\]

\[
\dot{\omega}_{(\mu)} + \frac{2}{3} \Theta \omega_\mu + \frac{1}{2} \text{curl} A_\mu = \sigma_{\mu \nu} \omega^\nu, \quad (A\ 13)
\]

\[
\dot{\sigma}_{(\mu \nu)} + \frac{2}{3} \Theta \sigma_{\mu \nu} + E_{\mu \nu} - \frac{1}{2} \pi_{\mu \nu} - \tilde{\nabla}_{(\mu} A_{\nu)} = -\sigma_{\alpha (\mu} \sigma_{\nu)}^{\alpha} - \omega_{(\mu} \omega_{\nu)} + A_{(\mu} A_{\nu)}, \quad (A\ 14)
\]

\[
\dot{E}_{(\mu \nu)} + \Theta E_{\mu \nu} - \text{curl} H_{\mu \nu} + \frac{1}{2} (\rho + p) \sigma_{\mu \nu} + \frac{1}{2} \tilde{\pi}_{(\mu \nu)} + \frac{1}{6} \Theta \pi_{\mu \nu} + \frac{1}{2} \tilde{\nabla}_{(\mu} q_{\nu)} = -A_{(\mu} q_{\nu)} + 2 A^{\alpha} \epsilon_{\alpha \beta (\mu} H_{\nu)}^\beta + 3 \sigma_{\alpha (\mu} E_{\nu)}^\alpha - \omega^\alpha \epsilon_{\alpha \beta (\mu} E_{\nu)}^\beta - \frac{1}{2} \sigma^\alpha_{\alpha (\mu} \pi_{\nu)}^{\alpha} - \frac{1}{2} \omega^\alpha \epsilon_{\alpha \beta (\mu} \pi_{\nu)}^{\beta}, \quad (A\ 15)
\]

and

\[
\dot{H}_{(\mu \nu)} + \Theta H_{\mu \nu} + \text{curl} E_{\mu \nu} - \frac{1}{2} \text{curl} \pi_{\mu \nu} = 3 \sigma_{\alpha (\mu} H_{\nu)}^\alpha - \omega^\alpha \epsilon_{\alpha \beta (\mu} H_{\nu)}^\beta - 2 A^{\alpha} \epsilon_{\alpha \beta (\mu} E_{\nu)}^\alpha - \frac{3}{2} \omega_{(\mu} q_{\nu)} + \frac{1}{2} \sigma^\alpha_{\alpha (\mu} \epsilon_{\nu)}^{\beta} q^{\beta}. \quad (A\ 16)
\]

**Constraint**

\[
\tilde{\nabla}^{\mu} \omega_\mu = A^{\mu} \omega_\mu, \quad (A\ 17)
\]

\[
\tilde{\nabla}^{\nu} \sigma_{\mu \nu} - \text{curl} \omega_\mu - \frac{2}{3} \tilde{\nabla}_\mu \Theta + q_\mu = -2 \epsilon_{\mu \nu \alpha} \omega^\nu A^\alpha, \quad (A\ 18)
\]

\[
\text{curl} \sigma_{\mu \nu} + \tilde{\nabla}_{(\mu} \omega_{\nu)} - H_{\mu \nu} = -2 A_{(\mu} \omega_{\nu)}, \quad (A\ 19)
\]

\[
\tilde{\nabla}^{\nu} E_{\mu \nu} + \frac{1}{2} \tilde{\nabla}^{\nu} \pi_{\mu \nu} - \frac{1}{3} \tilde{\nabla}_\mu p + \frac{1}{3} \Theta q_\mu = \epsilon_{\mu \nu \alpha} \sigma^\nu_\beta H^{\alpha \beta} - 3 H_{\mu \nu} \omega^\nu + \frac{1}{2} \sigma_{\mu \nu} q^\nu - \frac{3}{2} \epsilon_{\mu \nu \alpha} \omega^\nu q^\alpha, \quad (A\ 20)
\]

and

\[
\tilde{\nabla}^{\nu} H_{\mu \nu} + \frac{1}{2} \text{curl} q_\mu - (\rho + p) \omega_\mu = -\epsilon_{\mu \nu \alpha} \sigma^\nu_\beta E^{\alpha \beta} - \frac{1}{2} \epsilon_{\mu \nu \alpha} \sigma^\nu_\beta \pi^{\alpha \beta} + 3 E_{\mu \nu} \omega^\nu - \frac{1}{2} \pi_{\mu \nu} \omega^\nu. \quad (A\ 21)
\]
The energy and momentum conservation equations are the evolution equations (A 10) and (A 12). The total dynamical quantities have contributions from all dynamically significant particle species. Thus

\[
T^{\mu \nu} = \sum_{I} T_{I}^{\mu \nu} = \rho u^{\mu} u^{\nu} + ph^{\mu \nu} + 2q^{(\mu} u^{\nu)} + \pi^{\mu \nu}
\]

and

\[
T_{I}^{\mu \nu} = \rho_{I}^{*} u^{\mu}_{I} u^{\nu}_{I} + p_{I}^{*} h^{\mu \nu}_{I} + 2q^{*(\mu}_{I} u^{\nu)}_{I} + \pi^{* \mu \nu}_{I},
\]

where \(I = r, n, b, c, A\) labels the species. The asterisk on the dynamical quantities \(\rho_{I}^{*}, \ldots\) denotes that these quantities are measured, not in the \(u^{\mu}\)-frame, but in the \(I\)-frame, whose 4-velocity is given by

\[
u_{I}^{\mu} = \gamma_{I}(u^{\mu} + v_{I}^{\mu}), \quad v_{I}^{\mu} u_{\mu} = 0 \quad \text{and} \quad \gamma_{I} = (1 - v_{I}^{2})^{-1/2}.
\]

The fully nonlinear equations for the \(I\) dynamical quantities as measured in the fundamental \(u^{\mu}\)-frame are

\[
\rho_{I} = \rho_{I}^{*} + \{\gamma_{I}^{2} v_{I}^{2}(\rho_{I}^{*} + p_{I}^{*}) + 2\gamma_{I} q_{I}^{* \mu} v_{I \mu} + \pi_{I}^{* \mu \nu} v_{I \mu} v_{I \nu}\}, \quad (A 24)
\]

\[
p_{I} = p_{I}^{*} + \frac{1}{3} \{\gamma_{I}^{2} v_{I}^{2}(\rho_{I}^{*} + p_{I}^{*}) + 2\gamma_{I} q_{I}^{* \mu} v_{I \mu} + \pi_{I}^{* \mu \nu} v_{I \mu} v_{I \nu}\}, \quad (A 25)
\]

\[
q_{I}^{\mu} = q_{I}^{* \mu} + (\rho_{I}^{*} + p_{I}^{*}) v_{I}^{\mu} + \{(\gamma_{I} - 1) q_{I}^{* \mu} v_{I \mu} + \gamma_{I} q_{I}^{* \mu} v_{I \mu} + \gamma_{I} q_{I}^{* \mu} v_{I \mu} v_{I \nu} \}
\]

\[
+ \gamma_{I}^{2} v_{I}^{2}(\rho_{I}^{*} + p_{I}^{*}) v_{I}^{\mu} + \pi_{I}^{* \mu \nu} v_{I \mu} + \pi_{I}^{* \mu \nu} v_{I \mu} + \pi_{I}^{* \mu \nu} v_{I \mu} v_{I \mu} \}
\quad (A 26)
\]

and

\[
\pi_{I}^{* \mu \nu} = \pi_{I}^{* \mu \nu} + \left\{-2u^{(\mu} \pi_{I}^{* \nu)} v_{I \alpha} + \pi_{I}^{* \mu \nu} v_{I \alpha} u^{\mu} u^{\nu} + 2\gamma_{I} v_{I}^{(\mu} q_{I}^{* \nu)} \right. \]

\[
\left. - 2\gamma_{I} v_{I}^{* \mu} v_{I \alpha} u^{(\mu} v_{I}^{\nu)} - \frac{1}{3} \pi_{I}^{* \mu \nu} v_{I \alpha} h^{\mu \nu} + \gamma_{I}^{2} \left(\rho_{I}^{*} + p_{I}^{*}\right) v_{I}^{(\mu} v_{I}^{\nu)} \right\}. \quad (A 27)
\]

The terms in braces are the nonlinear corrections that vanish in the standard perturbed FLRW case. The total dynamical quantities in equations (A 10)–(A 21) are given by

\[
\rho = \sum_{I} \rho_{I}, \quad p = \sum_{I} p_{I}, \quad q^{\mu} = \sum_{I} q_{I}^{* \mu} \quad \text{and} \quad \pi^{\mu \nu} = \sum_{I} \pi_{I}^{* \mu \nu}.
\]

Assuming that the species are non-interacting, they each separately obey the energy and momentum conservation equations (A 10) and (A 12),

\[
\dot{\rho}_{I} + (\rho_{I} + p_{I}) \Theta + \hat{\nabla}_{\mu} q_{I}^{\mu} = -2A_{\mu} q_{I}^{\mu} - \sigma_{\mu \nu} \pi_{I}^{* \mu \nu} \quad (A 29)
\]

and

\[
\dot{q}_{I}^{(\mu)} + \frac{4}{3} \Theta q_{I}^{\mu} + (\rho_{I} + p_{I}) A^{\mu} + \hat{\nabla}_{\nu} p_{I} + \hat{\nabla}_{\rho} \pi_{I}^{* \mu \nu} - A_{\nu} \pi_{I}^{* \mu \nu} = -\sigma_{\mu \nu} q_{I}^{\nu} + \epsilon_{\mu \alpha \omega} q_{I}^{\alpha} - A_{\nu} \pi_{I}^{* \mu \nu},
\]

where the \(I\)-quantities are given by equations (A 24)–(A 27).
The covariant kinetic theory description starts by splitting the photon 4-momentum as

\[ p^\mu = E(u^\mu + e^\mu), \quad e^\mu e_\mu = 1 \quad \text{and} \quad e^\mu u_\mu = 0. \] (A 31)

Here \( E = -u_\mu p^\mu \) is the energy and \( e^\mu = p^{(\mu)}/E \) is the direction, as measured by a co-moving fundamental observer. Then the photon distribution function is decomposed into covariant harmonics via the expansion

\[ f(x, p) = f(x, E, e) = F + F_\mu e^\mu + F_{\mu\nu} e^\mu e^\nu + \cdots = \sum_{\ell \geq 0} F_M(x, E) e^{(M_\ell)}, \] (A 32)

where \( M_\ell = \mu_1 \mu_2 \cdots \mu_\ell \) and \( e^{M_\ell} = e^{\mu_1} \cdots e^{\mu_\ell} \). The PSTF multi-poles \( F_M \) are a covariant alternative to the usual expansion in spherical harmonics. The energy–momentum tensor is \( T^{\mu\nu}(x) = \int p^\mu p^\nu f(x, p) d^3p \) so that the first three multi-poles define the radiation dynamical quantities (in the \( u^\mu \)-frame),

\[ \rho_\ell = 4\pi \int_0^\infty E^3 F dE, \quad q_\ell = \frac{4\pi}{3} \int_0^\infty E^3 F^\mu dE \quad \text{and} \quad \pi^\mu_\ell = \frac{8\pi}{15} \int_0^\infty E^3 F^{\mu\nu} dE. \] (A 33)

Thus \( \Pi = (1/4\pi) \rho_\ell, \Pi_\mu = (3/4\pi) q_\ell^\mu, \Pi^{\mu\nu} = (15/8\pi) \pi^{\mu\nu}_\ell \), where the brightness multi-poles are

\[ \Pi_{\mu_1 \cdots \mu_\ell} = \int E^3 F_{\mu_1 \cdots \mu_\ell} dE. \] (A 34)

The collisionless Boltzmann (or Liouville) equation is

\[ \frac{df}{dv} = p^\mu \frac{\partial f}{\partial x^\mu} - \Gamma^\mu_{\alpha\beta} p^\alpha p^\beta \frac{\partial f}{\partial p_\mu} = 0, \] (A 35)

where \( p^\mu = dx^\mu/dv \), and we neglect polarization. The covariant multi-poles of \( df/dv \) are given by

\[
\frac{1}{E} \left( \frac{df}{dv} \right)_{M_\ell} = \dot{F}_{(M_\ell)} - \frac{1}{3} \Theta F^{\mu}_{M_\ell} + \tilde{\nabla}_{(\mu_\ell} F_{(M_{\ell-1})}^{\mu_\ell)} + \left( \frac{\ell + 1}{2\ell + 3} \right) \tilde{\nabla}^\mu F_{\mu M_\ell} - \frac{(\ell + 1)}{(2\ell + 3)} E^{-(\ell + 1)} [E^{\ell + 2} F_{\mu M_\ell}] Y A^\mu - E^{\ell} [E^{1 - \ell} F_{(M_{\ell-1})} \gamma_{M_\ell}] A_{M_\ell}
\]

\[
- \ell \omega^p \varepsilon_{\mu_\ell (M_{\ell-1})}^{\alpha} \frac{(\ell + 1)(\ell + 2)}{(2\ell + 3)(2\ell + 5)} E^{-(\ell + 2)} [E^{\ell + 3} F_{\nu M_\ell}] \sigma^{\mu\nu}
\]

\[
- \frac{2\ell}{(2\ell + 3)} E^{-1/2} [E^{3/2} F_{\nu (M_{\ell-1})}] \sigma_{\mu_\ell} \gamma^\nu - E^{\ell - 1} [E^{2 - \ell} F_{(M_{\ell-2})}] \sigma_{\mu_{\ell-1} \mu_\ell},
\] (A 36)

where a prime denotes \( \partial/\partial E \). This is a fully nonlinear expression.
Multiplying by $E^3$ and integrating over all energies leads to the brightness multi-pole evolution equations

$$0 = \dot{H} + \frac{4}{3} \Theta \Pi M + \tilde{\nabla} \langle \mu \Pi M_{e-1} \rangle + \frac{(\ell + 1)}{(2\ell + 3)} \tilde{\nabla} \Pi \ell M \ell + \frac{(\ell + 1)(\ell - 2)}{(2\ell + 3)} A^\ell \Pi \ell M \ell + (\ell + 3) A_{\mu \Pi M_{e-1}}$$

$$- \ell \omega^\alpha \tilde{\epsilon}_{\mu \alpha} \Pi M_{e-1} - \frac{(\ell - 1)(\ell + 1)(\ell + 2)}{(2\ell + 3)(2\ell + 5)} \sigma^\alpha \Pi \ell M \ell \ell$$

$$+ \frac{5\ell}{(2\ell + 3)} \sigma^\alpha \langle \mu \Pi M_{e-1} \rangle - (\ell + 2) \sigma_{\mu \ell M_{e-1}}.$$

(A 37)

The monopole evolution equation is just the energy conservation equation, i.e. equation (A 29) with $I = r$, the dipole is the momentum conservation equation (A 30) with $I = r$, and the quadrupole is (3.9).

References

Is the Universe homogeneous?


