arXiv:astro-ph/9503125 v1 2 Apr 1995

COURSE 1

COSMOLOGICAL DYNAMICS

EDMUND BERTSCHINGER

Department of Physics
MIT, Room 6-207
Cambridge, MA 02139, USA.

© *Elsevier Science Publishers B.V., 2002*

photograph of lecturer

Contents

0. Preface

The theory of large-scale structure is presently one of the most active research areas in cosmology. The important questions being studied include: Did structure form by gravitational instability? What are the nature and amount of dark matter? What is the background cosmological model? What were the initial conditions for structure formation? It is exciting that we can ask these questions seriously, knowing that observational tests are rapidly improving.

Numerous papers and reviews discuss specific theoretical models of large-scale structure, or specific theoretical techniques for constructing and analyzing models. However, there are few coherent presentations of the basic physical theory of the dynamics of matter and spacetime in cosmology. Although there are now several textbooks in this area, I think there is still room for further pedagogical development. My aim in these lecture notes is to provide a detailed yet readable introduction to cosmological dynamics.

Although I gave an evening seminar on N-body techniques for simulating large-scale structure, for reasons of length I have excluded that subject from these notes. The subject is presented elsewhere (e.g., Hockney & Eastwood 1981, Efstathiou et al. 1985, Bertschinger & Gelb 1991, and S. White's notes in this volume). Otherwise, these notes generally follow the lectures I gave in Les Houches, except that my lecture on Lagrangian fluid dynamics has been subsumed into the section on relativistic perturbation theory. The former subject is still evolving, and does not seem to be as fundamental as the subjects of my other lectures.

I would like to thank Andrew Hamilton, Lam Hui, Bhuvnesh Jain, Chung-Pei Ma, Dominik Schwarz, Uroš Seljak, and Simon White for useful comments and discussion, and Rennan Bar-Kana, Chung-Pei Ma, Nick Gnedin, and Marie Machacek for correcting several errors in early drafts. I am grateful to the organizers and students of the Les Houches Summer School for providing the opportunity to present this material. I appreciate the hospitality of John Bahcall and the Institute for Advanced Study, where much of the writing was done. This work was supported by NASA grants NAGW-2807 and NAG5-2816.

1. Elementary mechanics

This lecture applies elementary mechanics to an expanding universe. Attention is given to puzzles such as the role of boundary conditions and conservation laws.

1.1. Newtonian dynamics in cosmology

For a finite, self-gravitating set of mass points with positions $\boldsymbol{r}_i(t)$ in an otherwise empty universe, Newton's laws (assuming nonrelativistic motions and no non-gravitational forces) are

$$\frac{d^2\boldsymbol{r}_i}{dt^2} = \boldsymbol{g}_i \ , \quad \boldsymbol{g}_i = -\sum_{j\neq i} Gm_j \frac{(\boldsymbol{r}_i - \boldsymbol{r}_j)}{|\boldsymbol{r}_i - \boldsymbol{r}_j|^3} \ . \tag{1.1}$$

In the limit of infinitely many particles each with infinitesimal mass ρd^3r, we can also obtain $\boldsymbol{g}_i = \boldsymbol{g}(\boldsymbol{r}_i, t)$ as the irrotational solution to the Poisson equation,

$$\boldsymbol{\nabla}\cdot\boldsymbol{g} = -4\pi G\rho(\boldsymbol{r},t) \ , \quad \boldsymbol{\nabla}\times\boldsymbol{g} = 0 \ , \tag{1.2}$$

which may be written

$$\boldsymbol{g}(\boldsymbol{r},t) = -\int G\rho(\boldsymbol{r}',t)\frac{(\boldsymbol{r}-\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|^3}\, d^3r' \ . \tag{1.3}$$

The Newtonian potential ϕ, defined so that $\boldsymbol{g} = -\partial\phi/\partial\boldsymbol{r}$ (using partial derivatives to indicate the gradient with respect to $\boldsymbol{r}$), obeys $\boldsymbol{\nabla}^2\phi = 4\pi G\rho$.

If the mass density ρ is finite and nonzero only in a finite volume, then $\boldsymbol{g}$ (and also ϕ) generally converges to a finite value everywhere, with $g \to 0$ as $r \to \infty$. If, however, ρ remains finite as $r \to \infty$, then ϕ diverges and $\boldsymbol{g}$ depends on boundary conditions at infinity.

Consider the dilemma faced by Newton in his correspondence with Bentley concerning the gravitational field in cosmology (Munitz 1957). What is $\boldsymbol{g}$ in an infinite homogeneous medium? If we consider first a bounded sphere of radius R, Gauss' theorem quickly gives us $\boldsymbol{g} = -(4\pi/3)G\rho\boldsymbol{r}$ for $r < R$. This result is unchanged as $R \to \infty$, so we might conclude that $\boldsymbol{g}$ is well-defined at any finite r. Suppose, however, that the surface bounding the mass is a spheroid (a flattened or elongated sphere, whose cross-section is an ellipse) of eccentricity $e > 0$. In this case the gravity field is nonradial (see Binney & Tremaine 1987, §2.3, for expressions). The only difference in the mass distribution is in the shell between the spheroid and its circumscribed sphere, yet the gravity field is changed everywhere except at $\boldsymbol{r} = 0$. An inhomogeneous density field further changes $\boldsymbol{g}$. Thus, the gravity field in cosmology depends on boundary conditions at infinity.

There is an additional paradox of Newtonian gravity in an infinite homogeneous medium: $\boldsymbol{g} = 0$ at one point but is nonzero elsewhere (at least in the spherical and spheroidal examples given above), in apparent violation of the Newtonian relativity of absolute space. Newton avoided this problem (incorrectly, in hindsight) by assuming that gravitational forces due to mass at infinity cancel everywhere so that a static solution exists.

These problems are resolved in general relativity (GR), which forces us to complicate the treatment of Newtonian gravity in absolute space. First, in GR distant matter curves spacetime so that $(\boldsymbol{r}, t)$ do not provide good coordinates in cosmology. Second, in GR we must specify a global spacetime geometry explicitly taking into account distant boundary conditions.

What coordinates shall we take in cosmology? First note that a homogeneous self-gravitating mass distribution cannot remain static (unless non-Newtonian physics such as a fine-tuned cosmological constant is added to the model, as was proposed by Einstein in 1917). The observed mass distribution is (on average) expanding on large scales. For a uniform expansion, all separations scale in proportion with a cosmic scale factor $a(t)$. Even though the expansion is not perfectly uniform, it is perfectly reasonable to factor out the mean expansion to account for the dominant motions at large distances as in Figure 1. We do this by defining comoving coordinates $\boldsymbol{x}$ and conformal time τ as follows:

$$\boldsymbol{x} = \boldsymbol{r}/a(t) \ , \quad d\tau = dt/a(t) \ \text{ or } \ \tau = \int_0^t \frac{dt'}{a(t')} \ . \tag{1.4}$$

The starting time for the expansion is $\tau = 0$ and $t = 0$ when $a = 0$; if this time was nonexistent (or ill-defined in classical terms) then we can set the lower limit of integration for $\tau(t)$ to any convenient value. Although the units of a are arbitrary, I follow the standard convention of Peebles (1980) in setting $a = 1$ today when $t = t_0$ and $\tau = \tau_0$. A radiation source emitting radiation at $\tau < \tau_0$ has redshift $\Delta\lambda/\lambda_0 = z = -1 + a^{-1}$ where λ_0 is the rest wavelength.

For a perfectly uniform expansion, the comoving position vectors $\boldsymbol{x}$ remain fixed for all particles. For a perturbed expansion, each particle follows a trajectory $\boldsymbol{x}(\tau)$ [or $\boldsymbol{x}(t)$]. The comoving coordinate velocity, known also

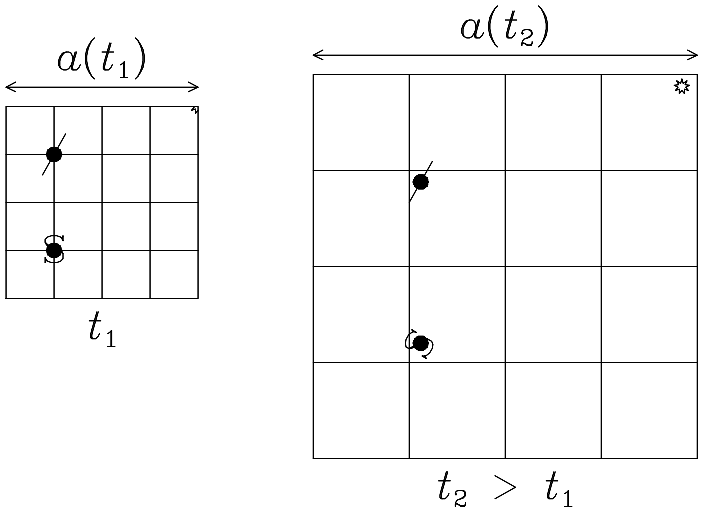

Fig. 1. Perturbed Hubble expansion.

as the peculiar velocity, is

$$\boldsymbol{v} \equiv \frac{d\boldsymbol{x}}{d\tau} = \frac{d\boldsymbol{r}}{dt} - H(t)\boldsymbol{r} \ , \tag{1.5}$$

where $H(t) = d\ln a/dt = a^{-2}da/d\tau$ is the Hubble parameter. Note that $\boldsymbol{v}$ is the proper velocity measured by a comoving observer at $\boldsymbol{x}$, i.e., one whose comoving position is fixed.

[The distinction between "proper" and "comoving" quantities is important. Proper quantities are physical observables, and they do not change if the expansion factor is multiplied by a constant. Thus, $\boldsymbol{v} = d\boldsymbol{x}/d\tau = (ad\boldsymbol{x})/(adt)$ is a proper quantity, while $d\boldsymbol{x}/dt$ is not. This is why I prefer τ rather than t as the independent variable.]

We shall assume that peculiar velocities are of the same order at all distances and in all directions, consistent with the choice of a homogeneous and isotropic mean expansion scale factor. These assumptions are consistent with the **Cosmological Principle**, which states that the universe is approximately homogeneous and isotropic when averaged over large volumes. In general relativity theory, the Cosmological Principle is applied by assuming that we live in a perturbed Robertson-Walker spacetime. Locally, the GR description is equivalent to Newtonian cosmology plus the boundary conditions that the mass distribution is (to sufficient accuracy) homogeneous and isotropic at infinity.

Unless otherwise stated, in this and the following lectures (until section 4) I shall use 3-vectors for spatial vectors assuming an orthonormal basis. Thus, $\boldsymbol{A}\cdot\boldsymbol{B} = A_iB_i = A^iB_i = A^iB^i$ with summation implied from $i=1$ to 3. Note that $A_i = A^i$ are Cartesian components, whether comoving or proper, and they are to be regarded (in this Newtonian treatment) as 3-vectors, not the spatial parts of 4-vectors. (If we were to use 4-vectors, then $A_i = g_{ij}A^j = a^2A^i$ in a Robertson-Walker spacetime. Because we are not using 4-vectors, there is no factor of a^2 distinguishing covariant and contravariant components.) This treatment requires space to be Euclidean, which is believed to be an excellent approximation everywhere except very near relativistic compact objects such as black holes and, possibly, on scales comparable to or larger than the Hubble distance c/H. (In section 4 the restrictions to Cartesian components and Euclidean space will be dropped.) Also, gradients and time derivatives will be taken with respect to the comoving coordinates: $\boldsymbol{\nabla}\equiv\partial/\partial\boldsymbol{x}$, $\dot{}\equiv\partial/\partial\tau$.

Before proceeding further we must derive the laws governing the mean expansion. Consider a spherical uniform mass distribution with mass density $\bar\rho$ and radius $r = xa(t)$ with $x=$ constant. Newtonian energy conservation states

$$\frac{1}{2}\left(\frac{dr}{dt}\right)^2 - \frac{GM}{r} = E\ ,$$

implying

$$\left(\frac{d\ln a}{d\tau}\right)^2 = (aH)^2 = \frac{8\pi}{3}Ga^2\bar\rho - K\ ,\quad K\equiv -2Ex^{-2}\ . \tag{1.6}$$

This result, known as the Friedmann equation, is valid (from GR) even if $\bar\rho$ includes relativistic particles or vacuum energy density $\rho_{\rm vac} = \Lambda/(8\pi G)$ (where Λ is the cosmological constant). The cosmic density parameter is $\Omega\equiv 8\pi G\bar\rho/(3H^2)$, so the Friedmann equation may also be written $K = (\Omega-1)(aH)^2$. Homogeneous expansion, with $a = a(\tau)$ independent of $\boldsymbol{x}$, requires $K=$ constant in addition to $\boldsymbol{\nabla}\bar\rho = 0$. In GR one finds that K is related to the curvature of space (i.e., of hypersurfaces of constant τ). The solutions of eq. (1.6) for zero-pressure (Friedmann) models, two-component models with nonrelativistic matter and radiation, and other simple equations of state may be found in textbooks (e.g., Padmanabhan 1993, Peebles 1993) or derived as good practice for the student.

At last we are ready to describe the motion of a nonuniform medium in Newtonian cosmology with mass density $\rho(\boldsymbol{x},\tau) = \bar\rho(\tau) + \delta\rho(\boldsymbol{x},\tau)$. We start from Newton's law in proper coordinates, $d^2\boldsymbol{r}/dt^2 = \boldsymbol{g}$, and transform to comoving coordinates and conformal time:

$$\frac{d^2\boldsymbol{x}}{d\tau^2} + \frac{\dot a}{a}\,\frac{d\boldsymbol{x}}{d\tau} + \boldsymbol{x}\frac{d}{d\tau}\left(\frac{\dot a}{a}\right) = -Ga^2\int(\bar\rho + \delta\rho)\frac{(\boldsymbol{x}-\boldsymbol{x}')}{|\boldsymbol{x}-\boldsymbol{x}'|^3}\,d^3x' \ .$$

We eliminate the homogeneous terms (those present in a homogeneous universe) as follows. First, assuming that the universe is, on average, spherically symmetric at large distance, the first term on the right-hand side becomes (from Gauss' theorem) $-(4\pi/3)Ga^2\bar\rho\boldsymbol{x}$. (This is where the boundary conditions at infinity explicitly are used.) To get the term proportional to $\boldsymbol{x}$ on the left-hand side, differentiate the Friedmann equation: $(\dot a/a)d(\dot a/a)/d\tau = (4\pi G/3)d(\bar\rho a^2)/d\tau$. For nonrelativistic matter, $\bar\rho \propto a^{-3}$, implying $d(\bar\rho a^2)/d\tau = -\dot a\bar\rho a$, so $d(\dot a/a)/d\tau = -(4\pi/3)Ga^2\bar\rho$. (If $\bar\rho$ includes relativistic matter, not only is $d\rho/d\tau$ changed, so is the gravitational field. Our derivation gives essentially the correct final result in this case, but its justification requires GR.) We conclude that the homogeneous terms cancel, so that the equation of motion becomes

$$\frac{d^2\boldsymbol{x}}{d\tau^2} + \frac{\dot a}{a}\,\frac{d\boldsymbol{x}}{d\tau} = -Ga^2\int\delta\rho(\boldsymbol{x}',\tau)\frac{(\boldsymbol{x}-\boldsymbol{x}')}{|\boldsymbol{x}-\boldsymbol{x}'|^3}\,d^3x' \equiv -\boldsymbol{\nabla}\phi' \ ,$$

where

$$\phi'(\boldsymbol{x},\tau) = -Ga^2\int\frac{\delta\rho(\boldsymbol{x}',\tau)\,d^3x'}{|\boldsymbol{x}-\boldsymbol{x}'|} \ .$$

Note that ϕ' is a proper quantity: $a^2d^3x'/|\boldsymbol{x}-\boldsymbol{x}'| \sim d^3r/|\boldsymbol{r}-\boldsymbol{r}'|$.

If $\int\delta\rho\,d^3x \to 0$ when the integral is taken over all space — as happens if the density field approaches homogeneity and isotropy on large scales, with $\bar\rho$ being the volume-averaged density — then ϕ' is finite and well-defined (except, of course, on top of point masses, which we ignore by treating the density field as being continuous). Newton's dilemma is then resolved: we have no ambiguity in the equation of motion for $\boldsymbol{x}(\tau)$. We conclude that ϕ', sometimes called the "peculiar" gravitational potential, is the correct Newtonian potential in cosmology provided we work in comoving coordinates. Therefore we shall drop the prime and the quaint historical adjective "peculiar." In summary, the equations of motion become

$$\frac{d^2\boldsymbol{x}}{d\tau^2} + \frac{\dot a}{a}\,\frac{d\boldsymbol{x}}{d\tau} = -\boldsymbol{\nabla}\phi \ , \quad \boldsymbol{\nabla}^2\phi = 4\pi Ga^2\delta\rho(\boldsymbol{x},\tau) \ . \tag{1.7}$$

As we shall see in section 4, the same equations follow in the weak-field ($|\phi| \ll c^2$), slow-motion ($v^2 \ll c^2$) limit of GR for a perturbed Robertson-Walker spacetime. If Newton had pondered more carefully the role of boundary conditions at infinity, he might have invented modern theoretical cosmology!

1.2. Lagrangian and Hamiltonian formulations

The equations of Newtonian cosmology may be derived from Lagrangian and Hamiltonian formulations. The latter is particularly useful for treatments of phase space.

In the Lagrangian approach, one considers the trajectories $\boldsymbol{x}(\tau)$ and the action $S[\boldsymbol{x}(\tau)]$. From elementary mechanics (with proper coordinates and no cosmology, yet), $S = \int L\,dt$ with Lagrangian $L = T - W = \frac{1}{2}mv^2 - m\phi$ for a particle moving in a potential ϕ (T is the kinetic energy and W is the gravitational energy). We now write a similar expression in comoving coordinates, bearing in mind that the action must be a proper quantity:

$$S = \int L(\boldsymbol{x}, \dot{\boldsymbol{x}}, \tau)\, d\tau \ , \quad L = a\left(\frac{1}{2}mv^2 - m\phi\right) \ , \tag{1.8}$$

where $\dot{\boldsymbol{x}} = \boldsymbol{v}$ is the peculiar velocity. We will show that eq. (1.8) is the correct Lagrangian by showing that it leads to the correct equations of motion.

Equations of motion for the trajectories follow from Hamilton's principle: the action must be stationary under small variations of the trajectories with fixed endpoints. Thus, we write $\boldsymbol{x}(\tau) \to \boldsymbol{x}(\tau) + \delta\boldsymbol{x}(\tau)$, $d\boldsymbol{x}/d\tau \to d\boldsymbol{x}/d\tau + (d/d\tau)\delta\boldsymbol{x}(\tau)$. The change in the action is

$$\begin{aligned}\delta S &= \int_{\tau_1}^{\tau_2} \left[\frac{\partial L}{\partial \boldsymbol{x}} \cdot \delta\boldsymbol{x} + \frac{\partial L}{\partial \dot{\boldsymbol{x}}} \cdot \frac{d}{d\tau}(\delta\boldsymbol{x})\right] d\tau \\ &= \int_{\tau_1}^{\tau_2} \left[\frac{\partial L}{\partial \boldsymbol{x}} - \frac{d}{d\tau}\left(\frac{\partial L}{\partial \dot{\boldsymbol{x}}}\right)\right] \cdot \delta\boldsymbol{x}(\tau)\, d\tau \ ,\end{aligned}$$

where we have integrated by parts assuming $(\partial L/\partial \dot{\boldsymbol{x}}) \cdot \delta\boldsymbol{x} = 0$ at $\tau = \tau_1$ and τ_2. Applying Hamilton's principle, $\delta S = 0$, we obtain the Euler-Lagrange equation (it works in cosmology, too!):

$$\frac{d}{d\tau}\left(\frac{\partial L}{\partial \dot{\boldsymbol{x}}}\right) - \frac{\partial L}{\partial \boldsymbol{x}} = 0 \ . \tag{1.9}$$

The reader may verify that substituting L from eq. (1.8) yields the correct equation of motion (1.7).

It is straightforward to extend this derivation to a system of self-gravitating particles filling the universe. The Lagrangian is

$$L = a\left(\sum_i \frac{1}{2}m_i v_i^2 - W\right) \ , \tag{1.10}$$

where the total gravitational energy excludes the part arising from the mean density:

$$W = \frac{1}{2}\left(\sum_i m_i\phi_i - a^3\bar{\rho}\int \phi\, d^3x\right) \ ,$$

$$\phi_i = -\left(\sum_{j\neq i}\frac{Gm_j}{a|\boldsymbol{x}_i - \boldsymbol{x}_j|} - Ga^2\bar{\rho}\int \frac{d^3x'}{|\boldsymbol{x} - \boldsymbol{x}'|}\right) \ , \tag{1.11}$$

where the factor $\frac{1}{2}$ is introduced to avoid double-counting pairs of particles. For a continuous mass distribution we obtain

$$W = \frac{1}{2}\int \phi\,\delta\rho\, a^3 d^3x = -\frac{1}{2}Ga^5\int d^3x_1\int d^3x_2\,\frac{\delta\rho(\boldsymbol{x}_1,\tau)\,\delta\rho(\boldsymbol{x}_2,\tau)}{|\boldsymbol{x}_1 - \boldsymbol{x}_2|} \ . \tag{1.12}$$

In the Hamiltonian approach one considers the trajectories in the single-particle (6-dimensional) phase space, $\{\boldsymbol{x}(\tau), \boldsymbol{p}(\tau)\}$. The aim is to obtain coupled first-order equations of motion for $\boldsymbol{x}(\tau)$ and $\boldsymbol{p}(\tau)$, known as Hamilton's equations, instead of a single second-order equation for $\boldsymbol{x}(\tau)$.

The derivation of Hamilton's equations has several steps. First we need the canonical momentum conjugate to $\boldsymbol{x}$:

$$\boldsymbol{p} \equiv \frac{\partial L}{\partial \dot{\boldsymbol{x}}} = am\boldsymbol{v} = am\frac{d\boldsymbol{x}}{d\tau} \ . \tag{1.13}$$

Note that $\boldsymbol{p}$ is *not* the proper momentum measured by a comoving observer: $m\boldsymbol{v}$ is. In Hamiltonian mechanics, one must use the conjugate momentum and not the proper momentum.

The next step is to eliminate $d\boldsymbol{x}/d\tau$ from the Lagrangian in favor of $\boldsymbol{p}$. We then transform from the Lagrangian to a new quantity called the Hamiltonian, using a Legendre transformation:

$$L(\boldsymbol{x}, \dot{\boldsymbol{x}}, \tau) \to H(\boldsymbol{x}, \boldsymbol{p}, \tau) \equiv \boldsymbol{p}\cdot\dot{\boldsymbol{x}} - L \ . \tag{1.14}$$

Notice that we transform L to H and $\dot{\boldsymbol{x}}$ to $\boldsymbol{p}$ (the latter through eq. 1.13). Why do we perform these transformations? The answer is that now Hamilton's principle gives the desired equations of motion for the phase-space trajectory $\{\boldsymbol{x}(\tau), \boldsymbol{p}(\tau)\}$. In phase space, Hamilton's principle says that the action $S = \int L\, d\tau = \int(\boldsymbol{p}\cdot\dot{\boldsymbol{x}} - H)\, d\tau$ must be stationary under independent variations of all phase space coordinates: $\boldsymbol{x}(\tau) \to \boldsymbol{x}(\tau) + \delta\boldsymbol{x}(\tau)$ and $\boldsymbol{p}(\tau) \to \boldsymbol{p}(\tau) + \delta\boldsymbol{p}(\tau)$. As an exercise, the reader can show, using a method similar to the derivation of the Euler-Lagrange equation above,

$$\frac{d\boldsymbol{x}}{d\tau} = \frac{\partial H}{\partial \boldsymbol{p}} \ , \quad \frac{d\boldsymbol{p}}{d\tau} = -\frac{\partial H}{\partial \boldsymbol{x}} \ , \tag{1.15}$$

provided that $\boldsymbol{p}\cdot\delta\boldsymbol{x}=0$ at the endpoints of τ.

In our case, $H=p^2/(2am)+am\phi$ (getting the a's right requires using the Legendre transformation), yielding

$$\frac{d\boldsymbol{x}}{d\tau}=\frac{\boldsymbol{p}}{am}\,,\quad \frac{d\boldsymbol{p}}{d\tau}=-am\boldsymbol{\nabla}\phi\ . \tag{1.16}$$

These equations could be combined to yield eq. (1.7), but in the Hamiltonian approach we prefer to think of two coupled evolution equations. This is particularly useful when studying the evolution of a system in phase space, as we shall do in section 3 with hot dark matter.

1.3. Conservation of momentum and energy?

Are total momentum and energy conserved in cosmology? This is a nontrivial question because the canonical momentum and Hamiltonian differ from the proper momentum and energy.

Consider first the momentum of a particle in an unperturbed Robertson-Walker universe. With no perturbations, $\phi=0$ so that Hamilton's equation for $\boldsymbol{p}$ becomes $d\boldsymbol{p}/d\tau=-am\boldsymbol{\nabla}\phi=0$, implying that the canonical momentum $\boldsymbol{p}$ is conserved. But, the *proper* momentum $m\boldsymbol{v}=a^{-1}\boldsymbol{p}$ measured by a comoving observer decreases as a increases. What happened to momentum conservation?

The key point is that $\boldsymbol{v}=d\boldsymbol{x}/d\tau$ is measured using a non-inertial (expanding) coordinate system. Suppose, instead, that we choose $\boldsymbol{v}$ to be a proper velocity measured relative to some fixed origin. Momentum conservation then implies $\boldsymbol{v}=\text{constant}$ (if $\boldsymbol{\nabla}\phi=0$, as we assumed above). At $\tau=\tau_1$ and τ_2, the particle is at $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, respectively. Because $d\boldsymbol{x}/d\tau$ gives the proper velocity relative to a *comoving* observer at the particle's position, at τ_1 we have $d\boldsymbol{x}/d\tau=\boldsymbol{v}-(\dot a/a)_1\boldsymbol{x}_1$, while at τ_2, $d\boldsymbol{x}/d\tau=\boldsymbol{v}-(\dot a/a)_2\boldsymbol{x}_2$. (The proper velocity relative to the fixed origin is $\boldsymbol{v}$ in both cases, but the Hubble velocity at the particle's position — the velocity of a comoving observer — changes because the particle's position has changed.) Combining these, we find $[\dot{\boldsymbol{x}}(\tau_2)-\dot{\boldsymbol{x}}(\tau_1)]/(\tau_2-\tau_1)\approx-(\dot a/a)[\boldsymbol{x}(\tau_2)-\boldsymbol{x}(\tau_1)]/(\tau_2-\tau_1)+O(\tau_2-\tau_1)$ or, in the limit $\tau_2-\tau_1\to0$, $d^2\boldsymbol{x}/d\tau^2=-(\dot a/a)d\boldsymbol{x}/d\tau$. This is precisely our comoving equation of motion in the case $\boldsymbol{\nabla}\phi=0$. Thus, the "Hubble drag" term $(\dot a/a)d\boldsymbol{x}/d\tau$ is merely a "fictitious force" arising from the use of non-inertial coordinates. Stated more physically, the particle appears to slow down because it is continually overtaking faster moving observers.

Energy conservation is more interesting. Let us check whether the Hamiltonian $H(\boldsymbol{x},\boldsymbol{p},\tau)$ is conserved. Using Hamilton's equations for a single particle, we get

$$\frac{dH}{d\tau} = \frac{\partial H}{\partial \boldsymbol{x}} \cdot \frac{d\boldsymbol{x}}{d\tau} + \frac{\partial H}{\partial \boldsymbol{p}} \cdot \frac{d\boldsymbol{p}}{d\tau} + \frac{\partial H}{\partial \tau} = \frac{\partial H}{\partial \tau} \ . \tag{1.17}$$

Using $H = p^2/(2am) + am\phi$, we obtain $dH/d\tau = -(\dot{a}/a)(p^2/2am) + md(a\phi)/d\tau$ which is nonzero even if $d\phi/d\tau = 0$. Is this lack of energy conservation due to the use of non-inertial coordinates? While the appearance of a Hubble-drag term may suggest this is the case, if we wish to obtain the total Hamiltonian (or energy) for a system of particles filling all of space, we have no choice but to use comoving coordinates.

Perhaps the Hamiltonian is not conserved because it is not the proper energy. To examine this possibility, we use the Hamiltonian for a system of particles in comoving coordinates, with $H = a(T + W)$. The proper kinetic energy (with momenta measured relative to comoving observers) is

$$T = \sum_i \frac{1}{2} m_i v_i^2 = \sum_i \frac{1}{2} \frac{p_i^2}{a^2 m_i} \ , \tag{1.18}$$

while the gravitational energy W is given in eq. (1.11). Holding fixed the momenta, we see that $a^2 T$ is a constant, implying $\partial(aT)/\partial\tau = -\dot{a}T$. Similarly, holding fixed the particle positions, we find that $a\phi$ is a constant, implying $\partial(aW)/\partial\tau = 0$. We thus obtain the Layzer-Irvine equation (Layzer 1963, Irvine 1965)

$$\frac{d}{d\tau}(T + W) = -\frac{\dot{a}}{a}(2T + W) \ . \tag{1.19}$$

Total energy (expressed in comoving coordinates) is not conserved in Newtonian cosmology. (This is also the case in GR — indeed, there is generally no unique scalar for the total energy in GR.) However, if almost all of the mass is in virialized systems obeying the classical virial theorem $2T + W \approx 0$, we recover approximate total energy conservation.

2. Eulerian fluid dynamics

2.1. Cosmological fluid equations

A fluid is a dense set of particles treated as a continuum. If particle collisions are rapid enough to establish a local thermal equilibrium (e.g., Maxwell-Boltzmann velocity distribution), the fluid is an ideal collisional gas. If collisions do not occur (e.g., a gas of dark matter particles), the gas is called collisionless. (I exclude incompressible fluids, i.e., liquids, from consideration because the gases considered in cosmology are generally very dilute and compressible.) The fluid equations discussed in this lecture apply only for a collisional gas (or a pressureless collisionless gas). They apply, for example, to baryons (hydrogen and helium gas) after recombination, to cold dark matter before trajectories intersect ("cold dust"), and (with relativistic corrections) to the coupled photon-baryon fluid before recombination.

I shall assume a nonrelativistic gas and ignore bulk electric and magnetic forces. These are not difficult to add, but the essential physics of cosmological fluid dynamics does not require them.

The fluid equations consist of mass and momentum conservation laws and an equation of state. Mass conservation is represented by the **continuity equation**. In proper coordinates $(\boldsymbol{r}, t)$ this is

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial \boldsymbol{r}} \cdot (\rho \boldsymbol{v}) = 0 \ , \quad \boldsymbol{v} = \frac{d\boldsymbol{r}}{dt} \ . \tag{2.1}$$

We convert to comoving coordinates $\tau = \int dt/a(t)$, $\boldsymbol{x} = \boldsymbol{r}/a(t)$, being careful to transform the partial derivatives as follows: $\partial/\partial t = (\partial\tau/\partial t)\partial/\partial\tau + (\partial\boldsymbol{x}/\partial t)\cdot\partial/\partial\boldsymbol{x}$, $\partial/\partial\boldsymbol{r} = a^{-1}\partial/\partial\boldsymbol{x} \equiv a^{-1}\boldsymbol{\nabla}$. We also rewrite the density and velocity by factoring out the mean behavior:

$$\rho = \bar{\rho}(1+\delta) \ , \quad \frac{d\boldsymbol{r}}{dt} = H\boldsymbol{r} + \boldsymbol{v} \tag{2.2}$$

where $\boldsymbol{v} = d\boldsymbol{x}/d\tau$ is now the peculiar velocity. The reader may easily show that eq. (2.1) becomes

$$\frac{\partial \delta}{\partial \tau} + \boldsymbol{\nabla} \cdot [(1+\delta)\boldsymbol{v}] = 0 \ . \tag{2.3}$$

Momentum conservation for an ideal fluid is represented by the **Euler equation** (Landau & Lifshitz 1959). It is most simply obtained by adding the pressure-gradient force to the equation of motion for a freely-falling mass element, eq. (1.7). In comoving coordinates, we find

$$\frac{d\boldsymbol{v}}{d\tau} + \frac{\dot{a}}{a}\boldsymbol{v} = -\boldsymbol{\nabla}\phi - \frac{1}{\rho}\boldsymbol{\nabla}p \ . \tag{2.4}$$

The time derivative is taken along the fluid streamline and is known as the convective or Lagrangian time derivative:

$$\frac{d}{d\tau} = \frac{\partial}{\partial \tau} + \boldsymbol{v} \cdot \boldsymbol{\nabla} \ . \tag{2.5}$$

Closing the fluid equations requires an evolution equation for the pressure or some other thermodynamic variable. Perhaps the most natural is the entropy. For a collisional gas, thermodynamics implies an **equation of state** $p = p(\rho, S)$ where S is the specific entropy. For example, for an ideal nonrelativistic monatomic gas, for reversible changes we have

$$T dS = d\left(\frac{3}{2}\frac{p}{\rho}\right) + p d\left(\frac{1}{\rho}\right) \ , \tag{2.6}$$

which says that the heat input to a fluid element equals the change in thermal energy plus the pressure work done by the element, i.e., energy is conserved. Combining this with the ideal gas law $p = \rho k_{\rm B} T/\mu$ where μ is the mean molecular mass and $k_{\rm B}$ is the Boltzmann constant, we obtain

$$p(\rho, S) = \rho^{5/3} \exp\left(\frac{2}{3}\frac{\mu}{k_{\rm B}} S\right) \ . \tag{2.7}$$

The equation of state must be supplemented by an evolution equation for the specific entropy. Outside of shock waves, the entropy evolution equation is

$$T\frac{dS}{d\tau} = a(\Gamma - \Lambda) \ , \tag{2.8}$$

where Γ and Λ are, respectively, the proper specific heating and cooling rates (in erg g^{-1} s^{-1}). They are determined by microphysical processes such as radiative emission and absorption, cosmic ray heating, Compton processes, etc. For the simplest case, adiabatic evolution, $\Gamma = \Lambda = 0$. For a realistic non-ideal gas, it may be necessary to evolve the radiation field, the ionization fraction, and other variables specifying the equation of state.

The fluid equations are much harder to solve than Newton's laws for particles falling under gravity, for several reasons. First, they are nonlinear partial differential equations rather than a set of coupled ordinary differential equations. Second, shock waves (discontinuities in ρ, p, S, and $\boldsymbol{v}$) prevent intersection of fluid elements. These discontinuities must be resolved (on a computational mesh or otherwise) and followed stably and accurately. Finally, heating and cooling for realistic gases are complicated and can lead to large temperature or entropy gradients that are difficult to resolve. An example of the latter is the sun, whose temperature changes by about 15 million K in a distance that is minuscule compared with cosmological distance scales.

Computational fluid dynamics is a difficult art but is important for galaxy formation. I shall not summarize the numerical methods here but refer the reader instead to the literature (e.g., Sod 1985, Leveque 1992, Monaghan 1992, Bryan et al. 1994, Kang et al. 1994).

Some of the most important effects of gas pressure can be gleaned from linear perturbation theory, in which we linearize the fluid equations about the uniform solution for an unperturbed Robertson-Walker spacetime. This technique is useful for checking for gravitational and other linear instabilities. Moreover, the linearized fluid equations may provide a reasonable description of large-scale, small-amplitude fluctuations in the (dark+luminous) matter, even if structure is nonlinear on small scales. This is a common assumption in large-scale structure theory. It is supported reasonably well by numerical simulations.

Linearizing the continuity and Euler equations gives

$$\dot{\delta} + \boldsymbol{\nabla}\cdot\boldsymbol{v} \approx 0\ , \quad \dot{\boldsymbol{v}} + \frac{\dot{a}}{a}\boldsymbol{v} \approx -\boldsymbol{\nabla}\phi - \frac{1}{\bar{\rho}}\boldsymbol{\nabla}p\ , \tag{2.9}$$

where an overdot denotes $\partial/\partial\tau$. The pressure gradient may be obtained from the equation of state $p = p(\rho, S)$. For an ideal nonrelativistic monatomic gas,

$$\frac{1}{\bar{\rho}}\boldsymbol{\nabla}p = c_{\rm s}^2\boldsymbol{\nabla}\delta + \frac{2}{3}T\boldsymbol{\nabla}S\ , \quad c_{\rm s}^2 = \frac{5}{3}\frac{p}{\bar{\rho}}\ . \tag{2.10}$$

Finally, we must linearize the entropy evolution equation. If the time scale for entropy changes is long compared with the acoustic or gravitational time scales, eq. (2.8) becomes $dS/d\tau \approx 0$. For the small peculiar velocities of linear perturbation theory this reduces to $\dot{S} \approx 0$.

There are five fluid variables (ρ, S, and three components of $\boldsymbol{v}$), hence five linearly independent modes. The general linear perturbation is a linear combination of these, which we now proceed to examine.

2.2. Linear instability 1: isentropic fluctuations and Jeans criterion

We begin with some nomenclature from thermodynamics. **Isentropic** means $\boldsymbol{\nabla}S = 0$: the same entropy everywhere. **Adiabatic** means $dS/d\tau = 0$: the entropy of a given fluid element does not change. The two concepts are distinct. It is common in cosmology to say "adiabatic" when one means "isentropic." This usage is confusing and I shall adopt instead the standard terminology from thermodynamics.

Isentropic fluctuations are the natural outcome of quantum fluctuations during inflation followed by reheating: rapid particle interactions in thermal equilibrium eliminate entropy gradients. If $\boldsymbol{\nabla}S = 0$, the linearized fluid and gravitational field equations are

$$\dot{\delta} + \boldsymbol{\nabla}\cdot\boldsymbol{v} = 0\ , \quad \dot{\boldsymbol{v}} + \frac{\dot{a}}{a}\boldsymbol{v} = -\boldsymbol{\nabla}\phi - c_{\rm s}^2\boldsymbol{\nabla}\delta\ , \quad \boldsymbol{\nabla}^2\phi = 4\pi G\bar{\rho}a^2\delta\ . \tag{2.11}$$

Combining these gives a damped, driven acoustic wave equation for δ:

$$\ddot{\delta} + \frac{\dot{a}}{a}\dot{\delta} = 4\pi G\bar{\rho}a^2\delta + c_{\rm s}^2\boldsymbol{\nabla}^2\delta\ . \tag{2.12}$$

Aside from the Hubble damping and gravitational source terms, this equation is identical to what one would get for linear acoustic waves in a static medium.

To eliminate the spatial Laplacian we Fourier transform the wave equation. For one plane wave, $\delta(\boldsymbol{x},\tau) \to \delta(\boldsymbol{k},\tau)\exp(i\boldsymbol{k}\cdot\boldsymbol{x})$. The wave equation becomes

$$\ddot{\delta} + \frac{\dot{a}}{a}\dot{\delta} = \left(4\pi G\bar{\rho}a^2 - k^2c_{\mathrm{s}}^2\right)\delta \equiv \left(k_{\mathrm{J}}^2 - k^2\right)c_{\mathrm{s}}^2\,\delta \ , \tag{2.13}$$

where we have defined the comoving Jeans wavenumber,

$$k_{\mathrm{J}} \equiv \left(\frac{4\pi G\bar{\rho}a^2}{c_{\mathrm{s}}^2}\right)^{1/2} \ . \tag{2.14}$$

Neglecting Hubble damping (by setting $a = 1$), the time dependence of the solution to eq. (2.13) would be $\delta \propto \exp(-i\omega\tau)$, yielding a dispersion relation very similar to that for high-frequency waves in a plasma, but with an important sign difference because gravity is attractive:

$$\omega^2 = \omega_{\mathrm{p}}^2 + k^2c^2 \quad \to \quad \omega^2 = -\omega_{\mathrm{J}}^2 + k^2c_{\mathrm{s}}^2 \ . \tag{2.15}$$

The plasma frequency is $\omega_{\mathrm{p}} = (4\pi n_e e^2/m_e)^{1/2}$ while the Jeans frequency is $\omega_{\mathrm{J}} = k_{\mathrm{J}}c_{\mathrm{s}} = (4\pi G\bar{\rho})^{1/2}$. Whereas electromagnetic waves with $\omega^2 < \omega_{\mathrm{p}}^2$ do not propagate ($k^2 < 0$ implies they are evanescent, e.g., they reflect off the Earth's ionosphere), gravitational modes with $k < k_{\mathrm{J}}$ are *unstable* ($\omega^2 < 0$), as was first noted by Jeans (1902). In physical terms, pressure forces cannot prevent gravitational collapse when the sound-crossing time λ/c_{s} is longer than the gravitational dynamical time $(G\rho)^{-1/2}$ for a perturbation of proper wavelength $\lambda = 2\pi a/k$.

Including the Hubble damping term slows the growth of the Jeans instability from exponential to a power of time for $k \ll k_{\mathrm{J}}$. In general there is one growing and one decaying solution for $\delta(k,\tau)$; these are denoted $\delta_{\pm}(k,\tau)$. For $c_{\mathrm{s}}^2 = 0$ and an Einstein-de Sitter (flat, matter-dominated) background with $a(\tau) \propto \tau^2$, $\delta_+ \propto \tau^2$ and $\delta_- \propto \tau^{-3}$. For $k \gg k_{\mathrm{J}}$, we obtain acoustic oscillations. In a static universe the acoustic amplitude for an adiabatic plane wave remains constant; in the expanding case it damps in general. An important exception is oscillations in the photon-baryon fluid in the radiation-dominated era; the amplitude of these oscillations is constant. (Showing this requires generalizing the fluid equations to a relativistic gas, a good exercise for the student.) In any case, acoustic oscillations suppress the growth relative to the long-wavelength limit.

It is interesting to write the linear wave equation in terms of ϕ rather than δ using $\nabla^2\phi = 4\pi G a^2\bar{\rho}\delta \propto a^{-1}\delta$ for nonrelativistic matter (with $c_s^2 \ll c^2$):

$$\ddot{\phi} + 3\frac{\dot{a}}{a}\dot{\phi} + \left(\frac{\ddot{a}}{a} - \frac{1}{2}\frac{\dot{a}^2}{a^2} - \frac{3}{2}K\right)\phi + k^2 c_s^2 \phi = 0 \ , \tag{2.16}$$

where we used the Friedmann equation (1.6); recall that $K = (\Omega - 1)(aH)^2$ is the spatial curvature constant. In a matter-dominated universe, differentiating the Friedmann equation gives $\ddot{a}/a - (1/2)\dot{a}^2/a^2 = -(1/2)K$, yielding

$$\ddot{\phi} + 3\frac{\dot{a}}{a}\dot{\phi} + \left(k^2 c_s^2 - 2K\right)\phi = 0 \ . \tag{2.17}$$

When written in terms of the gravitational potential rather than the density, the wave equation loses its gravitational source term.

The solutions to eq. (2.17) depend on the time-dependence of the sound speed as well as on the background cosmology. To get a rough idea of the behavior, consider the evolution of the potential in an Einstein-de Sitter universe filled with an ideal gas. For a constant sound speed, the solutions are

$$\phi_+(k,\tau) = \tau^{-2} j_2(kc_s\tau) \ , \ \ \phi_-(k,\tau) = \tau^{-2} y_2(kc_s\tau) \ , \ \ c_s = \text{const.} \ , \tag{2.18}$$

where j_2 and y_2 are the spherical Bessel functions of the first and second kinds of order 2. Although simple, this is not a realistic solution even before recombination (in that case, the photons and baryons behave as a single tightly-coupled relativistic gas, and relativistic corrections to the fluid equations must be added), except insofar as it illustrates the generic behavior of the two solutions: (damped) oscillations for $kc_s\tau \gg 1$ and power-law behavior for $kc_s\tau \ll 1$.

An alternative approximation, valid after recombination, is to assume that the baryon temperature roughly equals the photon temperature (this is a reasonable approximation because the small residual ionization thermally couples the two fluids for a long time even though there is negligible momentum transfer), $c_s^2 = c_{0s}^2 a^{-1}$ where c_{0s} is a constant. In this case the solutions are powers of τ:

$$\phi_\pm(k,\tau) = \tau^n \ , \ \ n = \frac{-5 \pm \sqrt{25 - 4(kc_{0s}\tau_0)^2}}{2} \ , \ \ T_{\text{gas}} \propto a^{-1} \ . \tag{2.19}$$

The solutions oscillate for $kc_s\tau_0 > 5/2$ and they damp for $kc_s\tau_0 > 0$.

In both of our solutions, and indeed for any reasonable equation of state in an Einstein-de Sitter universe, long-wavelength ($kc_s\tau \ll 1$) growing density modes have corresponding potential $\phi_+ =$ constant, while the decaying density modes have $\phi_- \propto \int a^{-3} d\tau$. The density perturbation and potential differ by a factor of $\bar{\rho}a^2 \propto a^{-1}$ from the Poisson equation. If $K < 0$ or $k^2 c_{\rm s}^2 > 0$, then ϕ_+ decays with time, although δ_+ still grows. Note that the important physical length scale where the **transfer function** $\phi_+(k,\tau)/\phi(k,0)$ falls significantly below unity is the *acoustic* comoving horizon distance $c_{\rm s}\tau$, not the causal horizon distance $c\tau$ or the Hubble distance c/H. Setting $c_{\rm s}$ to the acoustic speed of the coupled photon-baryon fluid at matter-radiation equality gives the physical scale at which the bend occurs, $c_{\rm s}\tau_{\rm eq}$, in the power spectrum of the standard cold dark matter and other models.

2.3. Linear instability 2: entropy fluctuations and isocurvature mode

Entropy gradients act as a source term for density perturbation growth. Using eq. (2.10) and repeating the derivation of the linear acoustic equation, we obtain (for $c_{\rm s}^2 \ll c^2$)

$$\ddot{\delta} + \frac{\dot{a}}{a}\dot{\delta} - 4\pi G\bar{\rho}a^2\delta - c_{\rm s}^2\boldsymbol{\nabla}^2\delta = \frac{2}{3}T\boldsymbol{\nabla}^2 S \ . \tag{2.20}$$

For *adiabatic* evolution, $\dot{S} = 0$, so what counts is the initial entropy gradient. Entropy gradients may be produced in the early universe by first-order phase transitions resulting in spatial variations in the photon/baryon ratio or other abundance ratios. If there were no entropy gradients present before such a phase transition, then the entropy variations can only have been produced by nonadiabatic processes. (This may explain the "adiabatic vs. isocurvature" nomenclature used by some cosmologists.) In practice, these entropy fluctuations are taken as initial conditions for subsequent adiabatic evolution.

Equation (2.20) is not applicable to the early universe because it assumes the matter is a one-component nonrelativistic gas. However, the behavior of its solutions are qualitatively similar to those for a relativistic multicomponent gas and so its analysis is instructive.

The **isocurvature mode** is given by the particular solution of density perturbation growth having $\delta = \dot{\delta} = 0$ but $\boldsymbol{\nabla}^2 S \neq 0$ at some early initial time τ_i. The initial conditions may be regarded as a perturbation in the equation of state in an otherwise unperturbed Robertson-Walker (constant spatial curvature) spacetime, accounting for the name "isocurvature." Variations in entropy at constant density correspond to variations in pressure, which lead through adiabatic expansion to changes in the density. Therefore, initial entropy fluctuations seed density fluctuations.

The solution to eq. (2.20) is obtained easily in Fourier space using the source-free (isentropic) solutions $\delta_\pm(k,\tau)$:

$$\delta_S(k,\tau) = -\frac{2}{3}k^2 S(k)\left[\;\delta_+(k,\tau)\int_{\tau_i}^{\tau} a'T'\delta'_-\,d\tau' \right.$$
$$\left. -\;\delta_-(k,\tau)\int_{\tau_i}^{\tau} a'T'\delta'_+\,d\tau'\right]\,, \qquad (2.21)$$

where primes are used to indicate that the variables are evaluated at $\tau=\tau'$. We see that both growing and decaying density perturbations are induced. After the source $(aT\delta_-)$ becomes small, the density fluctuations evolve the same way as isentropic fluctuations — e.g., they oscillate as acoustic waves if $kc_s\tau \gg 1$. To reinforce the point about nomenclature made earlier, I note that in our approximation, *both* isocurvature and "adiabatic" (i.e., isentropic) modes are *adiabatic* in the sense of thermodynamics: $\dot S = 0$ after the initial moment. For a realistic multi-component gas the evolution is not truly adiabatic, but that is a complication we shall not consider further. In the literature, modes are described as being adiabatic or isocurvature depending only on whether the initial density is perturbed with negligible initial entropy perturbation, or vice versa.

2.4. Vorticity — or potential flow?

With the growing and decaying isentropic perturbations, and the isocurvature mode, we have accounted for three of the expected five linear modes. The remaining two degrees of freedom were lost when we took the divergence of the Euler equation, thereby annihilating any transverse (rotational) contribution to $\boldsymbol{v}$. We consider them now.

Theorem: Any differentiable vector field $\boldsymbol{v}(\boldsymbol{x})$ may be written as a sum of longitudinal (curl-free) and transverse (divergence-free) parts, $\boldsymbol{v}_{||}$ and $\boldsymbol{v}_\perp$, respectively:

$$\boldsymbol{v}(\boldsymbol{x}) = \boldsymbol{v}_{||}(\boldsymbol{x}) + \boldsymbol{v}_\perp(\boldsymbol{x})\;,\quad \boldsymbol{\nabla}\times\boldsymbol{v}_{||} = \boldsymbol{\nabla}\cdot\boldsymbol{v}_\perp = 0\;. \qquad (2.22)$$

The proof follows by construction, by solving $\boldsymbol{\nabla}\cdot\boldsymbol{v}_{||}=\theta$ and $\boldsymbol{\nabla}\times\boldsymbol{v}_\perp=\omega$ where $\theta\equiv\boldsymbol{\nabla}\cdot\boldsymbol{v}$ and $\omega\equiv\boldsymbol{\nabla}\times\boldsymbol{v}$. In a flat Euclidean space, solutions are given by

$$\boldsymbol{v}_{||}(\boldsymbol{x}) = \frac{1}{4\pi}\int \theta(\boldsymbol{x}')\frac{(\boldsymbol{x}-\boldsymbol{x}')}{|\boldsymbol{x}-\boldsymbol{x}'|^3}\,d^3x'\;,\quad \theta(\boldsymbol{x})\equiv\boldsymbol{\nabla}\cdot\boldsymbol{v}\;, \qquad (2.23)$$

$$\boldsymbol{v}_\perp(\boldsymbol{x}) = \frac{1}{4\pi}\int \omega(\boldsymbol{x}')\times\frac{(\boldsymbol{x}-\boldsymbol{x}')}{|\boldsymbol{x}-\boldsymbol{x}'|^3}\,d^3x'\;,\quad \omega(\boldsymbol{x})\equiv\boldsymbol{\nabla}\times\boldsymbol{v}\;. \qquad (2.24)$$

Note that this decomposition is not unique; we may always add to $\boldsymbol{v}_{||}$ a curl-free solution of $\boldsymbol{\nabla}\cdot\boldsymbol{v}_{||}=0$ and to $\boldsymbol{v}_\perp$ a divergence-free solution of $\boldsymbol{\nabla}\times\boldsymbol{v}_\perp=0$ (e.g., constant vectors). With suitable boundary conditions (e.g., $\int \boldsymbol{v}_{||}\,d^3x=0$ when integrated over all space) this freedom can be eliminated. The variables θ and ω are called the (comoving) expansion scalar and vorticity vector, respectively.

In our preceding discussion of perturbation evolution we have implicitly considered only $\boldsymbol{v}_{||}$. The remaining two degrees of freedom correspond to the components of $\boldsymbol{v}_\perp$ (the transversality condition $\boldsymbol{\nabla}\cdot\boldsymbol{v}_\perp=0$ removes one degree of freedom from this 3-vector field). Fortunately, we can get a simple nonlinear equation for $\boldsymbol{v}_\perp$ — actually, for its curl, ω — by taking the curl of the Euler equation:

$$\begin{aligned}\dot{\omega}+\frac{\dot{a}}{a}\omega &= \boldsymbol{\nabla}\times(\boldsymbol{v}\times\omega)+\rho^{-2}(\boldsymbol{\nabla}\rho)\times(\boldsymbol{\nabla}p)\\ &= \boldsymbol{\nabla}\times(\boldsymbol{v}\times\omega)+\tfrac{2}{3}T(\boldsymbol{\nabla}\ln\rho)\times(\boldsymbol{\nabla}S)\end{aligned}\tag{2.25}$$

where we have assumed an ideal monatomic gas in writing the second form. The term arising from entropy gradients is called the **baroclinic** term. It is very important for the dynamics of the Earth's atmosphere and oceans (Pedolsky 1987).

An important general result follows from eq. (2.25), the **Kelvin Circulation Theorem**: If $\omega=0$ everywhere initially, then ω remains zero (even in the nonlinear regime) if the baroclinic term vanishes. (We are assuming that other torques such as magnetic ones vanish too.) The reason for the importance of this result in cosmology is that many models assume irrotational, isentropic initial conditions. With adiabatic evolution, it follows that $\omega=0$. Such a flow is also called potential flow because the velocity field may then be obtained from a velocity potential: $\boldsymbol{v}=\boldsymbol{v}_{||}=-\boldsymbol{\nabla}\Phi_v$.

Nonadiabatic processes (heating and cooling) and oblique shock waves can generate vorticity. In a collisionless fluid, if the fluid velocity is defined as the mass-weighted average of all the mass elements at a point, this averaging behaves like entropy production in regions where trajectories intersect, and so vorticity can be generated in the mean (fluid) velocity field. Vorticity also arises from isocurvature initial conditions. Equation (2.21) implies $\delta_S\propto\boldsymbol{\nabla}^2S$ for long wavelengths in the linear regime, giving a baroclinic torque proportional to $\boldsymbol{\nabla}\delta_S\times\boldsymbol{\nabla}S\propto\boldsymbol{\nabla}(\boldsymbol{\nabla}^2S)\times\boldsymbol{\nabla}S$, which is nonzero in general (though it appears only in second-order perturbation theory).

For most structure formation models, vorticity generation is quite small until shocks form (or trajectories intersect, for collisionless dark matter). In this case, one may obtain the velocity potential from the line integral of the velocity field:

$$\Phi_v(\boldsymbol{x}) = \Phi_v(0) - \int_0^{\boldsymbol{x}} \boldsymbol{v} \cdot d\boldsymbol{l} \ . \tag{2.26}$$

Taking the path to be radial with the observer in the middle allows one to reconstruct the velocity potential, and therefore the transverse velocity components, from the radial component. This idea underlies the potential flow reconstruction method, POTENT (Bertschinger & Dekel 1989). If the (smoothed) density fluctuations are sufficiently small for linear theory to be valid, we can estimate the density fluctuation field from an additional divergence. If pressure is unimportant, so $k \ll k_{\rm J}$ and $\delta \propto \delta_+(\tau)$, the linearized continuity equation gives

$$\boldsymbol{\nabla} \cdot \boldsymbol{v} = \theta = -\dot{\delta} = -aH \left(\frac{d \ln \delta_+}{d \ln a} \right) \delta \ . \tag{2.27}$$

For a wide range of cosmological models, $d\ln\delta_+/d\ln a \equiv f(\Omega) \approx \Omega^{0.6}$ depends primarily on the mass density parameter and weakly on other cosmological parameters (Peebles 1980, Lahav et al. 1991). Thus, combining measurements of $\boldsymbol{v}$ (radial components from galaxy redshifts and distances) and independent measurements of δ (from the galaxy density field plus an assumption about how dark matter is distributed relative to galaxies) allows estimation of Ω (Dekel et al. 1993). A review of the POTENT techniques and results is given by Dekel (1994).

3. Hot dark matter

The previous lecture studied the evolution of an ideal collisional gas including gravity and pressure. A gas of neutrinos, or of collisionless dark matter particles, behaves differently. In this lecture we investigate the evolution of a nonrelativistic collisionless gas whose particles have significant thermal speeds. (Relativistic kinetic theory is discussed by Stewart 1971, Bond & Szalay 1983, and Ma & Bertschinger 1994b.) An example is the gas of relic thermal neutrinos that decoupled at a temperature $k_{\rm B}T \sim 1$ MeV in the early universe. The present number density of these neutrinos (about 113 cm^{-3} for each of the three flavors) is such that a single massive type contributes $m_\nu c^2/(93\,h^2\,{\rm eV})$ to Ω, where $h = H_0/(100\,{\rm km\,s^{-1}\,Mpc^{-1}})$. Massive neutrinos are called hot dark matter because their thermal speeds significantly affect the gravitational growth of perturbations.

Before working out the detailed equations of motion for hot dark matter, it is useful to consider in general terms the effect of a thermal distribution.

Suppose we have a cold gas with no thermal motions. In this case it doesn't matter whether the gas is collisional or collisionless: gravitational instability amplifies the growing mode of irrotational density perturbations. What happens when we add thermal motions? We know the answer for a collisional gas: pressure stabilizes collapse for wavelengths less than the Jeans length, the distance sound waves travel in one gravitational dynamical time. For collisionless particles we also expect suppression. However, a collisionless gas cannot support sound waves, because no restoring force is provided by particle collisions.

A perfect collisional gas is fully described by its mass (or energy) density, fluid velocity, and temperature as functions of position. All other properties follow from the fact that the phase space density distribution is (locally) the thermal equilibrium distribution, e.g. Maxwell-Boltzmann. This is not true for a collisionless gas, whose complete description requires specifying the full phase space density.

For a collisionless gas, the velocity distribution function may be far from Maxwellian, so that the spatial stress tensor is not the simple diagonal form appropriate for an ideal gas. Instead there may be significant off-diagonal terms contributing **shear stress** that acts like viscosity in a weakly collisional fluid: it damps relative motions. We expect perturbations in a collisionless gas to be damped for wavelengths shorter than the distance traveled by particles with the characteristic thermal speed during one gravitational collapse time, the collisionless analogue of the Jeans length. Stated simply, overdense or underdense perturbations decay because the particles fly away from them at thermal speeds. This collisionless damping process is called free-streaming damping.

The characteristic thermal speed of massive neutrinos after they become nonrelativistic is

$$v_{\rm th} = \frac{k_{\rm B}T_\nu}{m_\nu c} = 50.4(1+z)\,(m_\nu c^2/{\rm eV})^{-1}\,{\rm km\,s^{-1}} \tag{3.1}$$

where we have used the standard big bang prediction $T_\nu = (4/11)^{1/3}\,T_\gamma$ (e.g., Kolb & Turner 1990) with $T_\gamma \approx 2.735$ K today. Multiplying $v_{\rm th}$ by the gravitational time $(4\pi G\bar\rho a^2)^{-1/2}$ gives the comoving free-streaming distance,

$$\lambda_{\rm fs} = 0.41\,(\Omega h^2)^{-1/2}\,(1+z)^{1/2}\,(m_\nu c^2/{\rm eV})^{-1}\,{\rm Mpc}\ . \tag{3.2}$$

At any time, fluctuations with wavelength less than about $\lambda_{\rm fs}$ are damped; much longer wavelength fluctuations grow with negligible suppression.

The free-streaming distance does not really grow without bound as $z \to \infty$ because the neutrino thermal speed cannot exceed c. Applying this limit gives a maximum comoving free-streaming distance of

$$\lambda_{\mathrm{fs,max}} = 31.8\,(\Omega h^2)^{-1/2}\,(m_\nu c^2/\mathrm{eV})^{-1/2}\,\mathrm{Mpc}\ . \tag{3.3}$$

Thus, unless they are regenerated by perturbations in other components (as happens, for example, in a model with hot and cold dark matter), primeval density fluctuations in massive neutrinos with wavelength smaller than this rather large scale will be erased by free-streaming damping. A more quantitative treatment is presented below using the actual evolution equations for the neutrino phase space density distribution.

3.1. Tremaine-Gunn bound

Before treating the phase space evolution, we discuss another important consequence of finite neutrino thermal speed: high-speed neutrinos cannot be tightly packed into galaxy halos. This fact can be used to place a lower bound on the neutrino mass if neutrinos make up the dark matter in galaxy halos (Tremaine & Gunn 1979).

The initial phase space density for massive neutrinos is a relativistic Fermi-Dirac distribution (preserved from the time when the neutrinos decoupled in the early universe):

$$f = \frac{2h_{\mathrm{P}}^{-3}}{\exp(pc/k_{\mathrm{B}}T_0)+1} \equiv f_0(\boldsymbol{p})\ , \tag{3.4}$$

where $\boldsymbol{p}$ is the comoving canonical momentum of eq. (1.13), h_{P} is Planck's constant (with a subscript to distinguish it from the scaled Hubble constant), and $T_0 = aT_\nu$ is the present neutrino "temperature." The decrease of T_ν with time is compensated for by the factor a relating proper momentum to comoving momentum. Ignoring perturbations, the present-day distribution for massive neutrinos is the relativistic Fermi-Dirac — not the equilibrium nonrelativistic distribution — because the phase space distribution was preserved after neutrino decoupling.

Tremaine & Gunn (1979) noted that because of phase mixing (discussed further below), the maximum coarse-grained phase space density of massive neutrinos today is less than the maximum of $f_0(\boldsymbol{p})$, h_{P}^{-3}. If massive neutrinos dominate the mass in galactic halos, this must be no less than the phase space density needed for self-gravitating equilibrium. This bound can be used to set a lower limit on the neutrino mass if one assumes that the neutrinos constitute the halo dark matter.

Although the neutrino mass bound is somewhat model-dependent because the actual coarse-grained distribution in galactic halos is unknown, we can get a reasonable estimate by assuming an isothermal sphere: a

Maxwell-Boltzmann distribution with constant velocity dispersion σ^2 (at $a = 1$ so that there is no distinction between proper and comoving):

$$f(\boldsymbol{r},\boldsymbol{p}) = (2\pi m_\nu^2\sigma^2)^{-3/2} n(r)\, \exp\left(\frac{-p^2}{2m_\nu^2\sigma^2}\right) \ . \tag{3.5}$$

In a self-gravitating system there are a family of spherical density profiles $\rho(r) = m_\nu n(r)$ obeying hydrostatic equilibrium:

$$\frac{1}{\rho}\frac{dP}{dr} = -\frac{GM(<r)}{r^2} = -\frac{4\pi G}{r^2}\int_0^r r^2\rho(r)\, dr \ . \tag{3.6}$$

The simplest case is the singular isothermal sphere with $\rho \propto r^{-2}$; the reader can easily check that $\rho = \sigma^2/(2\pi G r^2)$. Imposing the phase space bound at radius r then gives

$$m_\nu > (2\pi)^{-5/8}\left(G h_{\rm P}^3 \sigma r^2\right)^{-1/4} \ . \tag{3.7}$$

Up to overall numerical factors, this is the Tremaine-Gunn bound.

The singular isothermal sphere is probably a good model where the rotation curve produced by the dark matter halo is flat, but certainly breaks down at small radius. Because the neutrino mass bound is stronger for smaller σr^2, the uncertainty in the halo core radius (interior to which the mass density saturates) limits the reliability of the neutrino mass bound.

For the Local Group dwarf galaxies in Draco and Ursa Minor, measurements of stellar velocity dispersions suggest σ is a few to about 10 km s^{-1} (Pryor & Kormendy 1990). If these galaxies have isothermal halos at $r = 1$ kpc, the crude bound of eq. (3.7) implies m_ν is greater than a few eV.

3.2. Vlasov equation

We now present a rigorous treatment of the evolution of perturbations in a nonrelativistic collisionless gas, based on the evolution of the phase space distribution. The single-particle phase space density $f(\boldsymbol{x},\boldsymbol{p},\tau)$ is defined so that $f d^3x d^3p$ is the number of particles in an infinitesimal phase space volume element. We shall use comoving spatial coordinates $\boldsymbol{x}$ and the associated conjugate momentum $\boldsymbol{p} = am\dot{\boldsymbol{x}}$ (eq. 1.13). Note that $d^3x d^3p = m^3 d^3r d^3v$ is a proper quantity so that f is the proper (physical) phase space density.

If the gas is perfectly collisionless, f obeys the Vlasov (or collisionless Boltzmann) equation of kinetic theory,

$$\frac{Df}{D\tau} \equiv \frac{\partial f}{\partial \tau} + \frac{d\boldsymbol{x}}{d\tau}\cdot\frac{\partial f}{\partial \boldsymbol{x}} + \frac{d\boldsymbol{p}}{d\tau}\cdot\frac{\partial f}{\partial \boldsymbol{p}} = 0 \ . \tag{3.8}$$

This equation expresses conservation of particles along the phase space trajectory $\{\boldsymbol{x}(\tau), \boldsymbol{p}(\tau)\}$. Using Hamilton's equations (1.16) for nonrelativistic particles, we obtain

$$\frac{\partial f}{\partial \tau} + \frac{\boldsymbol{p}}{am} \cdot \frac{\partial f}{\partial \boldsymbol{x}} - am \boldsymbol{\nabla} \phi \cdot \frac{\partial f}{\partial \boldsymbol{p}} = 0 \ . \tag{3.9}$$

The Vlasov equation is supposed to apply for the coarse-grained phase space density for a collisionless gas in the absence of two-body correlations (Ichimaru 1992). Often, however, the statistical assumptions underlying the use of the Vlasov equation are vague. To clarify its application we digress to present a derivation using the Klimontovich (1967) approach to kinetic theory.

Consider one realization of a universe filled with particles following phase space trajectories $\{\boldsymbol{x}_i(\tau), \boldsymbol{p}_i(\tau)\}$ (i labels the particles). The *exact* single-particle phase space density (called the Klimontovich density) is written by summing over Dirac delta functions:

$$f(\boldsymbol{x}, \boldsymbol{p}, \tau) = \sum_i \delta[\boldsymbol{x} - \boldsymbol{x}_i(\tau)]\, \delta[\boldsymbol{p} - \boldsymbol{p}_i(\tau)] \ . \tag{3.10}$$

No statistical averaging or coarse-graining has been applied; f is the fine-grained density for one universe. This phase space density obeys the Klimontovich (1967) equation, which is of exactly the same form as eq. (3.8). The proof follows straightforwardly from substituting eq. (3.10) into eq. (3.8).

The Klimontovich density retains all information about the microstate of a system because it specifies the trajectories of all particles. This is far too much information to be practical. We must reduce the information content by performing some averaging or coarse-graining. This averaging is taken over a statistical ensemble of microstates corresponding to a given macrostate — for example, microstates with the same phase space density averaged over small phase space volumes containing many particles on average. We denote the averages using angle brackets $\langle\rangle$, without being very precise about the ensemble adopted for the coarse-graining.

The discreteness effects of individual particles are accounted for by the s-particle distribution functions ($s = 1,\, 2,\, \ldots$) f_s, which are defined using a standard cluster expansion:

$$\langle f(\boldsymbol{x}, \boldsymbol{p}, \tau) \rangle = \left\langle \sum_i \delta(\boldsymbol{x} - \boldsymbol{x}_i)\, \delta(\boldsymbol{p} - \boldsymbol{p}_i) \right\rangle \equiv f_1(\boldsymbol{x}, \boldsymbol{p}, \tau) \ , \tag{3.11}$$

$$
\begin{aligned}
&\langle f(\boldsymbol{x}_1,\boldsymbol{p}_1,\tau)\, f(\boldsymbol{x}_2,\boldsymbol{p}_2,\tau)\rangle = \\
&\qquad \left\langle \sum_{i=j} \delta(\boldsymbol{x}_1-\boldsymbol{x}_i)\,\delta(\boldsymbol{p}_1-\boldsymbol{p}_i)\,\delta(\boldsymbol{x}_2-\boldsymbol{x}_i)\,\delta(\boldsymbol{p}_2-\boldsymbol{p}_i) \right\rangle + \\
&\qquad\quad \left\langle \sum_{i\neq j} \delta(\boldsymbol{x}_1-\boldsymbol{x}_i)\,\delta(\boldsymbol{p}_1-\boldsymbol{p}_i)\,\delta(\boldsymbol{x}_2-\boldsymbol{x}_j)\,\delta(\boldsymbol{p}_2-\boldsymbol{p}_j) \right\rangle \qquad (3.12)\\
&\quad = \delta(\boldsymbol{x}_1-\boldsymbol{x}_2)\delta(\boldsymbol{p}_1-\boldsymbol{p}_2) f_1(\boldsymbol{x}_1,\boldsymbol{p}_1,\tau) + f_2(\boldsymbol{x}_1,\boldsymbol{p}_1,\boldsymbol{x}_2,\boldsymbol{p}_2,\tau)\ ,
\end{aligned}
$$

and so on. We further write f_2 as a sum of uncorrelated and correlated parts,

$$
f_2(\boldsymbol{x}_1,\boldsymbol{p}_1,\boldsymbol{x}_2,\boldsymbol{p}_2,\tau) = f_1(\boldsymbol{x}_1,\boldsymbol{p}_1,\tau) f_1(\boldsymbol{x}_2,\boldsymbol{p}_2,\tau) + f_{2c}(\boldsymbol{x}_1,\boldsymbol{p}_1,\boldsymbol{x}_2,\boldsymbol{p}_2,\tau)\ . \quad (3.13)
$$

This equation defines f_{2c}, known in kinetic theory as the irreducible two-particle correlation function. If there are no pair correlations in phase space, $f_{2c}=0$.

We now ensemble-average the Klimontovich equation, recalling that it is identical to eq. (3.9) provided we use the Klimontovich density. If ϕ is a specified external potential, neglecting self-gravity, we see that f_1 obeys the Vlasov equation. However, if ϕ is computed self-consistently from the particles, the $m\boldsymbol{\nabla}\phi\cdot(\partial f/\partial\boldsymbol{p})$ term is quadratic in the Klimontovich density, yielding an additional correlation term from eqs. (3.12) and (3.13) after coarse-graining. This term is not present in the Vlasov equation.

The contribution to the gravity field from the particles is (cf. eq. 1.11)

$$
\begin{aligned}
-\boldsymbol{\nabla}\phi(\boldsymbol{x},\tau) = -\frac{Gm}{a}\int d^3x'\, d^3p'\, f(\boldsymbol{x}',\boldsymbol{p}',\tau)\frac{(\boldsymbol{x}-\boldsymbol{x}')}{|\boldsymbol{x}-\boldsymbol{x}'|^3} \\
+G\bar{\rho}a^2\int d^3x'\,\frac{(\boldsymbol{x}-\boldsymbol{x}')}{|\boldsymbol{x}-\boldsymbol{x}'|^3}\ , \qquad (3.14)
\end{aligned}
$$

where the second term, required in comoving coordinates, removes the contribution from the mean uniform background.

Combining our results now yields the exact kinetic equation for the one-particle phase space density f_1:

$$
\begin{aligned}
&\frac{\partial f_1}{\partial\tau} + \frac{\boldsymbol{p}}{am}\cdot\frac{\partial f_1}{\partial\boldsymbol{x}} - am\boldsymbol{\nabla}\phi\cdot\frac{\partial f_1}{\partial\boldsymbol{p}} = \\
&Gm^2\int d^3x' d^3p'\,\frac{(\boldsymbol{x}-\boldsymbol{x}')}{|\boldsymbol{x}-\boldsymbol{x}'|^3}\cdot\frac{\partial}{\partial\boldsymbol{p}} f_{2c}(\boldsymbol{x},\boldsymbol{p},\boldsymbol{x}',\boldsymbol{p}',\tau) \qquad (3.15)
\end{aligned}
$$

where $-\nabla\phi$ is given by eq. (3.14) using f_1 for f, and adding any other contribution from other sources. Equation (3.15) is called the first BBGKY hierarchy equation (Peebles 1980, Ichimaru 1992). It differs from the Vlasov equation by a correlation integral term.

If there are no phase space correlations, as would occur if we had a smooth collisionless fluid, then the one-particle or coarse-grained distribution obeys the Vlasov equation of kinetic theory. Correlations may be introduced by gravitational clustering, which couples f_{2c} to f_1. One may derive an evolution equation for f_{2c} — the second BBGKY hierarchy equation — by averaging $f\partial f/\partial\tau$, but it involves f_{3c}, and so on. The result is an infinite hierarchy of coupled kinetic equations, the BBGKY hierarchy.

For some cases, Boltzmann's hypothesis of molecular chaos may hold, implying $f_{2c} = 0$ except at binary collisions, with the right-hand side of eq. (3.15) becoming a Boltzmann collision operator. Fortunately, for the particles of interest here — neutrinos — the gravitational (and non-gravitational, after neutrino decoupling) collision time is so long that the correlation integral is completely negligible. Thus, hot dark matter composed of massive neutrinos obeys the Vlasov equation after decoupling. From now on we shall drop the subscript 1 from f.

We now return to our main line of development to discuss phase mixing. The Vlasov equation implies conservation of phase space density, but a given initial volume $d^3x d^3p$ evolves in a complicated way (i.e., the trajectories of particles initially inside this volume may be highly complicated). Consider the initial phase space element shown in Figure 2a, extracted from a one-dimensional N-body simulation. Figures 2b and 2c show the phase space distribution at a later time, with each particle's trajectory evolved according to Hamilton's equations without (Fig. 2b) and with (Fig. 2c) gravity, respectively. In both cases the area $dx\,dp$ of the phase space element is identical to the initial area as a consequence of the Vlasov equation.

Figure 2c illustrates the process known as phase mixing: the phase space structure becomes highly convoluted as particles make multiple orbits. Regions of initially high phase space density can end up entwined with regions of initially low phase space density. Although the density is conserved along each phase space trajectory, if the distribution is coarse-grained (averaged over finite phase space volume), the resulting coarse-grained density is not conserved. The maximum coarse-grained density can only decrease, as we noted previously in the discussion of the Tremaine-Gunn bound.

The process of phase-mixing is complicated, and the only practical means of integrating the Vlasov equation for such an evolved collisionless system is by N-body simulation: the phase space is sampled with discrete particles at some initial time and the particle trajectories are computed, providing

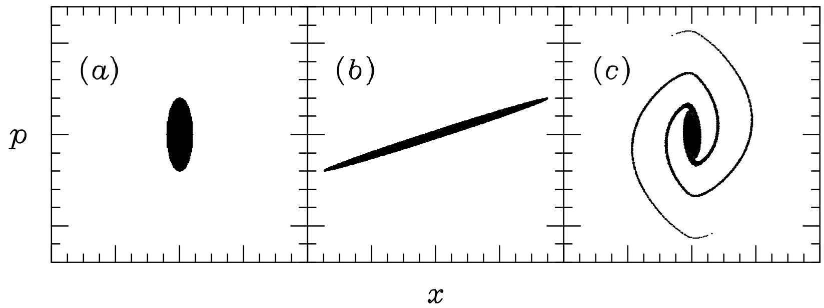

Fig. 2. Phase space evolution. (a) Initial conditions. (b) Evolved state without gravity. (c) Evolved state with gravity.

a sample of the evolved phase space. However, analytical methods can be used while the phase space distribution is only slightly perturbed from the homogeneous equilibrium distribution. These methods, presented in the next two subsections, will help us to understand free-streaming damping in detail.

3.3. Nonrelativistic evolution in an external gravitational field

In this section we consider hot dark matter made of nonrelativistic massive neutrinos with $\Omega_\nu \ll \Omega$ so that their self-gravity is unimportant. The gravitational potential $\phi(\boldsymbol{x}, \tau)$ (using comoving coordinates) is assumed to be given from other sources such as cold dark matter in a mixed hot and cold dark matter model.

We can solve the Vlasov equation (3.9) approximately by replacing $\partial f/\partial \boldsymbol{p}$ with the unperturbed term $\partial f_0/\partial \boldsymbol{p}$. This approximation is valid for $|f - f_0| \ll f_0$, and should suffice to demonstrate the collisionless damping of small-amplitude fluctuations.

A quadrature solution of the Vlasov equation can be obtained provided that we change the time variable from τ to $s = \int d\tau/a = \int dt/a^2$ and then Fourier transform the spatial variable:

$$f(\boldsymbol{x}, \boldsymbol{p}, \tau(s)) = \int d^3k\, e^{i\boldsymbol{k}\cdot\boldsymbol{x}}\, \hat{f}(\boldsymbol{k}, \boldsymbol{p}, s) \ . \tag{3.16}$$

The gravitational potential ϕ is transformed similarly. Integrating eq. (3.9) over s, we obtain the solution

$$\hat{f}(\boldsymbol{k},\boldsymbol{p},s) = \hat{f}(\boldsymbol{k},\boldsymbol{p},s_i)\, e^{-i\boldsymbol{k}\cdot\boldsymbol{u}(s-s_i)} + im\left(\boldsymbol{k}\cdot\frac{\partial f_0}{\partial \boldsymbol{p}}\right)\int_{s_i}^{s} ds'\, a^2(s')\hat{\phi}(\boldsymbol{k},s')\, e^{-i\boldsymbol{k}\cdot\boldsymbol{u}(s-s')} \ , \tag{3.17}$$

where $\boldsymbol{u} = \boldsymbol{p}/m$ and s_i is an initial time. If the initial phase space distribution is unperturbed, then $\hat{f}(\boldsymbol{k},\boldsymbol{p},s_i) = f_0(p)\,\delta(\boldsymbol{k})$. Note that the complex exponentials in eq. (3.17) correspond to the propagation of the phase space density along the characteristics $d\boldsymbol{x}/ds = \boldsymbol{u}$. This motion is called free-streaming.

To understand the behavior of the free-streaming solution, let us examine the integral term of eq. (3.17), which is proportional to

$$\int_0^{s-s_i} dy\, a^2(y+s_i)\,\hat{\phi}(\boldsymbol{k},y+s_i)\, e^{-i\beta y} \ , \tag{3.18}$$

where $\beta \equiv \boldsymbol{k}\cdot\boldsymbol{p}/m$ and $y = s' - s_i$. For sufficiently slowly moving neutrinos, β is small enough so that $\beta y \ll 1$. This condition corresponds to a free-streaming distance along $\boldsymbol{k}$ that is much less than k^{-1}. These neutrinos do not move far from the crests and troughs of the plane wave perturbation. Neglecting the exponential, the time dependence of the solution is the same as for cold dark matter.

If, however, $\beta y \gg 1$, corresponding to neutrinos traveling across many wavelengths of a perturbation, the rapid oscillations of the exponential lead to cancellation in the integrand of eq. (3.18) and suppression of the neutrino phase space density perturbation. This effect, known as free-streaming damping, occurs because neutrinos that are initially at the crests or troughs of density waves move so far that they distribute themselves almost uniformly. The small gravitational acceleration induced by the external potential is inadequate to collect the fast-moving neutrinos in dense regions.

Thus, perturbations can grow only for the neutrinos that move less than about one wavelength per Hubble time. Our analysis confirms the rough picture we sketched in the beginning of this lecture.

We can obtain the net density perturbation (in Fourier space) by integrating eq. (3.17) over momenta:

$$\begin{aligned}\frac{\hat{n}_\nu(\boldsymbol{k},s)}{n_0} &\equiv \frac{1}{n_0}\int d^3p\, \hat{f}(\boldsymbol{k},\boldsymbol{p},s) \\ &= \delta(\boldsymbol{k}) - k^2\int_{s_i}^{s} ds'\, a^2(s')\,\hat{\phi}(\boldsymbol{k},s')\,(s-s')\,F\left[\frac{k(s-s')}{m}\right] \ ,\end{aligned} \tag{3.19}$$

where $n_0 = \int d^3p\, f_0(p)$ is the mean comoving number density and F is the Fourier transform — with respect to the momentum! — of the unperturbed distribution function:

$$F(q) = \frac{1}{n_0} \int d^3p\, e^{-i\boldsymbol{p}\cdot\boldsymbol{q}}\, f_0(p)\ . \tag{3.20}$$

For the relativistic Fermi-Dirac distribution appropriate to hot dark matter, F has the series representation (Bertschinger & Watts 1988)

$$F(q) = \frac{4}{3\,\zeta(3)} \sum_{n=1}^{\infty} (-1)^{n+1} \frac{n}{(n^2 + q^2 p_0^2)^2}\ , \quad p_0 \equiv \frac{k_{\rm B} T_0}{c}\ , \tag{3.21}$$

where $\zeta(3) = 1.202\ldots$ is the Riemann zeta function and $F(0) = 1$.

Equation (3.19) does not give much insight into free-streaming damping. To get a better feel for the physics, as well as a simpler approximation for treating hot dark matter, we now show how to convert eq. (3.19) into a differential equation for the evolution of the hot dark matter density perturbation similar to eq. (2.12) for a perfect collisional fluid. This may seem impossible a priori — how can the dispersive behavior of a collisionless gas be represented by fluid-like differential equations? — but we shall see that it is possible if we approximate $f_0(p)$ by a form differing slightly from the Fermi-Dirac distribution. The results, although not exact, will give us additional insight into the behavior of collisionless damping.

The first step is to rewrite eq. (3.19) for the Fourier transform of the density fluctuation $\hat{\delta}_\nu$:

$$\hat{\delta}_\nu(\boldsymbol{k}, s) = -km \int_{s_i}^{s} ds'\, a^2(s')\, \hat{\phi}(\boldsymbol{k}, s')[qF(q)]\ , \quad q \equiv \frac{k(s-s')}{m}\ . \tag{3.22}$$

Next, we differentiate twice with respect to the time coordinate s:

$$\frac{\partial \hat{\delta}_\nu}{\partial s} = -k^2 \int_{s_i}^{s} ds'\, a^2(s')\, \hat{\phi}(\boldsymbol{k}, s')\, \frac{d}{dq}[qF(q)]\ , \tag{3.23}$$

$$\frac{\partial^2 \hat{\delta}_\nu}{\partial s^2} = -k^2 a^2(s) \hat{\phi}(\boldsymbol{k}, s) - \frac{k^3}{m} \int_{s_i}^{s} ds'\, a^2(s')\, \hat{\phi}(\boldsymbol{k}, s')\, \frac{d^2}{dq^2}[qF(q)]\ . \tag{3.24}$$

Note the appearance of a non-integrated source term in the second derivative, arising because $d(qF)/dq$ does not vanish at $s = s'$ $(q = 0)$ while qF does.

Next, we note that if $d^2(qF)/dq^2$ were to equal a linear combination of $d(qF)/dq$ and (qf), then we could write the integral in equation (3.24) as a linear combination of $\partial\hat{\delta}_\nu/\partial s$ and $\hat{\delta}_\nu$. Unfortunately, this is not the case for $F(q)$ given by eq. (3.21). However, it is true for the family of distribution functions whose Fourier transforms are

$$F_\gamma(q) = \exp(-\gamma q p_0) \ , \tag{3.25}$$

for any dimensionless constant γ. This defines the family of phase space density distributions

$$f_\gamma(p) = n_0 \int \frac{d^3q}{(2\pi)^3}\, e^{i\boldsymbol{p}\cdot\boldsymbol{q}}\, F_\gamma(q) = \frac{n_0}{\pi^2(\gamma p_0)^3}\left(1+\frac{p^2}{\gamma^2 p_0^2}\right)^{-2} \ . \tag{3.26}$$

For this form of unperturbed distribution we have

$$\frac{d^2}{dq^2}(qF_\gamma) = -2\gamma p_0 \frac{d}{dq}(qF_\gamma) - (\gamma p_0)^2 qF_\gamma \ . \tag{3.27}$$

Combining eqs. (3.22)–(3.24) and (3.27), we get

$$\frac{\partial^2\hat{\delta}_\nu}{\partial s^2} + 2\frac{\gamma p_0 k}{m}\frac{\partial\hat{\delta}_\nu}{\partial s} + \frac{\gamma^2 p_0^2 k^2}{m^2}\hat{\delta}_\nu = -k^2 a^2(s)\hat{\phi}(\boldsymbol{k}, s) \ . \tag{3.28}$$

To put this result into a form similar to the acoustic wave equation we derived for a collisional fluid, we define the characteristic proper thermal speed

$$c_\nu \equiv \gamma\frac{k_{\rm B}T_\nu}{mc} = \frac{\gamma p_0}{ma} \ . \tag{3.29}$$

Next, we change the time variable from s back to τ with $d\tau/ds = a$. Finally, we assume that the source term gravitational potential $\hat{\phi}$ is given by the Poisson equation for a perturbation $\delta_{\rm c}$ in a component with mean mass density $\bar{\rho}_{\rm c}$ (e.g., cold dark matter — recall that we are neglecting the self-gravity of the neutrinos). Dropping the hat on $\hat{\delta}_\nu$, the result is

$$\ddot{\delta}_\nu + \left(\frac{\dot{a}}{a} + 2kc_\nu\right)\dot{\delta}_\nu + k^2c_\nu^2\delta_\nu = 4\pi G a^2\bar{\rho}_{\rm c}\delta_{\rm c} \ . \tag{3.30}$$

This equation was first derived by Setayeshgar (1990). It is approximate (not exact) for the linear evolution of massive neutrinos because we replaced the Fermi-Dirac distribution by eq. (3.26). It is not difficult to show that eq. (3.26) is the only form of the distribution function for which eq. (3.17) can be reduced to a differential equation for $\delta_\nu(\boldsymbol{k}, \tau)$. (Even the Maxwell-Boltzmann distribution fails — a collisionless gas with this distribution initially does not evolve the same way as a collisional gas with the Maxwell-Boltzmann distribution function for all times.) One should also bear in mind that δ_ν does not contain all the information needed to characterize perturbations in a collisionless gas (Ma & Bertschinger 1994a). Complete information resides in $\hat{f}(\boldsymbol{k}, \boldsymbol{p}, s)$.

Even if eq. (3.30) is not exact for massive neutrinos and does not fully specify the perturbations, it provides an extremely helpful pedagogic guide to the physics of collisionless damping. We see at once that a gravitational source can induce density perturbations in a collisionless component, but the source competes agains acoustic $(k^2c_\nu^2)$ and damping $(\dot{a}/a+2kc_\nu)$ terms. Roughly speaking, hot dark matter behaves like a collisional gas with an extra free-streaming damping term.

Does the $k^2c_\nu^2$ term imply that a collisionless gas can support acoustic oscillations? To check this we consider the limit $kc_\nu\tau \gg 1$ so that the Hubble damping and gravitational source terms are negligible. We then have

$$\ddot{\delta}_\nu + 2\omega_\nu\dot{\delta}_\nu + \omega_\nu^2\delta_\nu \approx 0 \ , \quad \omega_\nu = kc_\nu \ . \tag{3.31}$$

Because ω_ν changes very slowly with time compared with the oscillation timescale ω^{-1}, eq. (3.31) is a linear differential equation with constant coefficients and is easily solved to give the two modes

$$\delta_\nu \propto \tau e^{-\omega_\nu\tau} \quad \text{or} \quad e^{-\omega_\nu\tau} \ , \quad \omega_\nu\tau \gg 1 \ . \tag{3.32}$$

Neither solution oscillates! The first one begins to grow but is rapidly damped on a timescale ω_ν^{-1}, after the typical neutrino has had time to cross one wavelength.

Because the damping time $(kc_\nu)^{-1}$ is proportional to the wavelength, short-wavelength perturbations are damped most strongly. At any given time τ, perturbations of comoving wavelength less than about $c_\nu\tau$ are attenuated. This is precisely the free-streaming distance we introduced in the beginning of this lecture, equation (3.2).

Our results enable us to understand why the hot dark matter transfer function is similar to that of cold dark matter for long wavelengths but cuts off sharply for short wavelengths (Bond & Szalay 1983). During the radiation-dominated era, $a(\tau) \propto \tau$. While the massive neutrinos were relativistic, $c_\nu \approx c$ was constant. The comoving free-streaming distance increased, $c_\nu\tau \propto a$, with hot dark matter perturbations being erased on scales up to the Hubble distance. After the neutrinos became nonrelativistic, however, c_ν is given by eq. (3.29), $c_\nu \propto a^{-1}$. Thus, the free-streaming distance saturates at the Hubble distance when the neutrinos become nonrelativistic. During the matter-dominated era, $a(\tau) \propto \tau^2$ (while $\Omega \approx 1$) so that the free-streaming distance decreases: $c_\nu \tau \propto a^{-1/2}$. However, free-streaming has already erased the hot dark matter perturbations on scales up to the maximum free-streaming distance, eq. (3.3). Only if the perturbations are re-seeded, e.g. by cold dark matter or topological defects, will small-scale power be restored to the hot dark matter.

3.4. Nonrelativistic evolution including self-gravity

Now that we have developed the basic techniques for solving the linearized nonrelativistic Vlasov equation, adding self-gravity of the collisionless particles is easy. We simply add a contribution to ϕ arising from δ_ν. In eq. (3.17), if we have a mixture of hot and cold dark matter, $\hat{\phi} \to (\hat{\phi}_{\rm c} + \hat{\phi}_\nu)$; additional contributions may be added as appropriate. Equation (3.22) becomes

$$\hat{\delta}_\nu(\boldsymbol{k}, s) = \frac{m}{k} \int_{s_i}^{s} ds'\, a^2(s')\, [qF(q)]\, 4\pi G a^2(s') \times \left[\bar{\rho}_{\rm c}(s')\hat{\delta}_{\rm c}(\boldsymbol{k}, s') + \bar{\rho}_\nu(s')\hat{\delta}_\nu(\boldsymbol{k}, s')\right] \ . \tag{3.33}$$

This equation was first derived (in a slightly different form) by Gilbert (1966) and is known as the Gilbert equation. Note that in the self-gravitating case δ_ν appears both inside and outside an integral. Equation (3.33) is a Volterra integral equation of the second kind. Bertschinger & Watts (1988) present a numerical quadrature solution method.

Using the same trick as in the previous subsection, we can convert the Gilbert equation to a differential equation for δ_ν, if the unperturbed phase space density distribution is approximated by the form $f_\gamma(p)$ of eq. (3.26). The result is

$$\ddot{\delta}_\nu + \left(\frac{\dot{a}}{a} + 2kc_\nu\right)\dot{\delta}_\nu + k^2 c_\nu^2 \delta_\nu = 4\pi G a^2 \left[\bar{\rho}_{\rm c}\delta_{\rm c} + \bar{\rho}_\nu \delta_\nu\right] \ . \tag{3.34}$$

With a suitable choice for the parameter γ, the solution of eq. (3.34) provides a good match (to within a few percent, in general) to the solution of the Gilbert equation using the correct Fermi-Dirac distribution for massive neutrinos (Setayeshgar 1990). Therefore, it may be used for obtaining quick estimates of the density perturbations of nonrelativistic hot dark matter.

4. Relativistic cosmological perturbation theory

4.1. Introduction

This section is an expanded version of my fifth lecture at Les Houches. One lecture gave barely enough time to introduce the essential ideas of relativistic perturbation theory: classification of metric perturbations, the linearized Einstein equations, and gauge modes. Understanding the physics of these topics, as well as the relativistic generalizations of my previous lectures, requires a much deeper immersion. Unable to find a pedagogical treatment in the existing literature that matches these needs to my satisfaction, I have developed the subject more fully in these written lecture notes. They are not a complete guide to relativistic perturbation theory but rather a starting point from which the reader may delve into the increasingly rich literature of applications. This section is self-contained and may be read independently of the previous sections, although the reader may find it interesting to contrast the nonrelativistic presentations of sections 1 and 2 with the relativistic treatment given below.

4.1.1. Synopsis

According to the Newtonian perspective of gravity and cosmology, spacetime is flat and absolute, gravity is action at a distance, and particle dynamics is given by Newton's second law $\boldsymbol{F} = m\boldsymbol{a}$ or, equivalently, by Hamilton's principle of least action. The Einsteinian perspective is quite different: spacetime is a curved manifold which evolves causally through the Einstein field equations in response to sources, and particle dynamics is given in absence of nongravitational forces by geodesic motion. In this section I attempt not only to present the essentials of relativistic gravitational dynamics, but also to show how it reduces to and extends Newtonian cosmology in the appropriate limit.

One of the main purposes of these notes is to provide a clear explanation of the scalar, vector, and tensor modes of gravitational perturbations. (We shall follow the customary usage in this subject by referring to different spatial symmetry components as "modes" even when they are not expanded in any basis eigenfunctions. Thus, the "scalar mode" is described, in part, by a field $\phi(x^\mu)$ that is a scalar under spatial coordinate transformations but is not restricted to being a single Fourier component or other harmonic basis function.) Newtonian gravity corresponds to the former (the scalar mode), while the latter (vector and tensor modes) represent the relativistic effects of gravitomagnetism and gravitational radiation, which have no counterpart in Newtonian gravity although they are similar to electromagnetic phenomena. If the motion of sources is expanded in powers of v/c, the vector and tensor gravitational fields are $O(v/c)$ and $O(v/c)^2$ times the Newtonian field, respectively. On terrestrial scales the vector and tensor modes are extremely weak — they have not been detected in the laboratory, although satellite experiments are planned to search for the former through the Lense-Thirring "gravitomagnetic moment" precession, and large interferometric detectors are being built to measure gravitational radiation — but they could have important consequences for the evolution of large-scale matter and radiation fluctuations, including the production of anisotropy in the microwave background radiation.

The Newtonian limit corresponds to weak gravitational fields (black holes are to be avoided) and slow motions ($v^2 \ll c^2$, for both sources and test particles). For nearly all cosmological applications it is sufficient to consider only weak fields — small perturbations of the spacetime metric around a homogeneous and isotropic background spacetime. At the same time it is usually safe to assume that the gravitational sources are nonrelativistic, although the test particles (e.g., photons) need not be. Because the weak-field, slow source motion limit does not necessarily imply small density fluctuations, we can (and will) investigate nonlinear particle and fluid dynamics even while treating the metric perturbations and source velocities as being small.

In sections 4.2–4.5 we shall develop the machinery for cosmological perturbation theory using the methods developed by Lifshitz, Peebles, Bardeen, Kodama & Sasaki, and others. We discuss the consequences of gauge invariance — the invariance of physical quantities to small changes in the spacetime coordinates — and summarize the standard results in the synchronous gauge of Lifshitz (1946).* In section 4.6 we introduce a new gauge that clarifies how general relativity extends Newtonian gravity in the weak-field limit and in section 4.7 we attempt to clarify the physical content of general relativity theory in this limit. In section 4.8 we shall see how simply and clearly the Hamiltonian formulation of particle dynamics follows from general relativity. Finally, in section 4.9 we introduce an alternative fully nonlinear formulation of general relativity due to Ehlers, Ellis and others, and we demonstrate its connection with the Lagrangian fluid dynamics that was discussed in my fourth lecture.

* Apparently it is not widely known that Lifshitz' paper is published in English and is available in many libraries. This classic paper was remarkably complete, including a full treatment of the scalar, vector, and tensor decomposition in open and closed universes and a concise solution to the gauge mode problem; it presented solutions for perfect fluids in matter- and radiation-dominated universes; and it contrasted isentropic (adiabatic) and entropy fluctuations.

We shall not discuss the relativistic Boltzmann equation nor the classification of isentropic and isocurvature initial conditions. In the nonrelativistic limit, these topics have already been covered in my preceding lectures. Neither shall we discuss the physics of microwave background anisotropy or the evolution of perturbations in specific models. Our aim here is to derive and comprehend the gravitational field equations, not their solution. Although this goal is restricted, we shall see that the physical content is sufficiently rich. After working through these notes the reader may wish to consult one of the many books or articles discussing the detailed evolution for a variety of models (e.g., Lifshitz & Khalatnikov 1963; Peebles & Yu 1970; Weinberg 1972; Peebles 1980; Press & Vishniac 1980; Wilson & Silk 1981; Wilson 1983; Bond & Szalay 1983; Zel'dovich & Novikov 1983; Kodama & Sasaki 1984, 1986; Efstathiou & Bond 1986; Bond & Efstathiou 1987; Ratra 1988; Holtzman 1989; Efstathiou 1990; Mukhanov, Feldman & Brandenberger 1992; Liddle & Lyth 1993; Peebles 1993; Ma & Bertschinger 1994b).

Understanding these notes will not require much experience with general relativity, although some background is helpful. The reader can test the waters by examining the following summary of essential general relativity and differential geometry. While some mathematical formalism is needed to get started, the focus thereafter will remain as much as possible on physics.

4.1.2. Summary of essential relativity

We adopt the following conventions and notations, similar to those of Misner, Thorne & Wheeler (1973). Units are chosen so that $c = 1$. The metric signature is $(-,+,+,+)$. The unperturbed background spacetime is Robertson-Walker with scale factor $a(\tau)$ expressed in terms of conformal time. A dot (or ∂_τ) indicates a conformal time derivative. The comoving expansion rate is written $\eta(\tau) \equiv \dot{a}/a = aH$. The scale factor obeys the Friedmann equation,

$$\eta^2 = \frac{8\pi}{3} G a^2 \bar{\rho} - K \ . \tag{4.1}$$

The Robertson-Walker line element is written in the general form using conformal time τ and comoving coordinates x^i:

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = a^2(\tau)\left[-d\tau^2 + \gamma_{ij}(x^k) dx^i dx^j\right] \ . \tag{4.2}$$

Latin indices (i, j, k, etc.) indicate spatial components while Greek indices (μ, ν, λ, etc.) indicate all four spacetime components; we assume a coordinate basis for tensors. Summation is implied by repeated upper and lower indices. The inverse 4-metric $g^{\mu\nu}$ (such that $g^{\mu\nu} g_{\nu\kappa} = \delta^\mu{}_\kappa$) is used to raise spacetime indices while the inverse 3-metric γ^{ij} ($\gamma^{ij}\gamma_{jk} = \delta^i{}_k$) is used to raise indices of 3-vectors and tensors. Three-tensors are defined in the spatial hypersurfaces of constant τ with metric γ_{ij} and they shall be clearly distinguished from the spatial components of 4-tensors. We shall see as we go along how this "3+1 splitting" of spacetime works when there are metric perturbations.

Many different spatial coordinate systems may be used to cover a uniform-curvature 3-space. For example, there exist quasi-Cartesian coordinates (x, y, z) in terms of which the 3-metric components are

$$\gamma_{ij} = \delta_{ij}\left[1 + \frac{K}{4}\left(x^2 + y^2 + z^2\right)\right]^{-2} \ . \tag{4.3}$$

We shall use 3-tensor notation to avoid restricting ourselves to any particular spatial coordinate system. Three-scalars, vectors, and tensors are invariant under transformations of the spatial coordinate system in the background spacetime (e.g., rotations). A 3-vector may be written $\boldsymbol{A} = A^i\boldsymbol{e}_i$ where $\boldsymbol{e}_i$ is a basis 3-vector obeying the dot product rule $\boldsymbol{e}_i \cdot \boldsymbol{e}_j = \gamma_{ij}$. A second-rank 3-tensor may be written (using dyadic notation and the tensor product) $\mathsf{h} = h^{ij}\boldsymbol{e}_i \otimes \boldsymbol{e}_j$. We write the spatial gradient 3-vector operator $\boldsymbol{\nabla} = \boldsymbol{e}^i\partial_i$ ($\partial_i \equiv \partial/\partial x^i$) where $\boldsymbol{e}^i \cdot \boldsymbol{e}_j = \delta^i{}_j$. The experts will recognize $\boldsymbol{e}^i$ as a basis one-form but we can treat it as a 3-vector $\boldsymbol{e}^i = \gamma^{ij}\boldsymbol{e}_j$ because of the isomorphism between vectors and one-forms. Because the basis 3-vectors in general have nonvanishing gradients, we define the covariant derivative (3-gradient) operator $\boldsymbol{\nabla}_i$ with $\boldsymbol{\nabla}_i\gamma_{jk} = 0$. If the space is flat ($K = 0$) and we use Cartesian coordinates, then $\gamma_{ij} = \delta_{ij}$, $\boldsymbol{\nabla}_i = \partial_i$, and the 3-tensor index notation reduces to elementary Cartesian notation. If $K \neq 0$, the 3-tensor equations will continue to look like those in flat space (that is why we use a 3+1 splitting of spacetime!) except that occasionally terms proportional to K will appear in our equations.

Our application is not restricted to a flat Robertson-Walker background but allows for nonzero spatial curvature. This complicates matters for two reasons. First, we cannot assume Cartesian coordinates. As a result, for example, the Laplacian of a scalar and the divergence and curl of a 3-vector involve the determinant of the spatial metric, $\gamma \equiv \det\{\gamma_{ij}\}$:

$$\begin{aligned} \boldsymbol{\nabla}^2\phi \equiv \gamma^{-1/2}\partial_i\left(\gamma^{1/2}\gamma^{ij}\partial_j\phi\right) \ , \quad \boldsymbol{\nabla}\cdot\boldsymbol{v} \equiv \gamma^{-1/2}\partial_i\left(\gamma^{1/2}v^i\right) \ , \\ \boldsymbol{\nabla}\times\boldsymbol{v} \equiv \epsilon^{ijk}(\partial_i v_j)\boldsymbol{e}_k \ , \end{aligned} \tag{4.4}$$

where $\epsilon^{ijk} = \gamma^{-1/2}[ijk]$ is the three-dimensional Levi-Civita tensor, with $[ijk] = +1$ if $\{ijk\}$ is an even permutation of $\{123\}$, $[ijk] = -1$ for an odd permutation, and 0 if any two indices are equal. The factor $\gamma^{-1/2}$

ensures that ϵ^{ijk} transforms like a tensor; as an exercise one can show that $\epsilon_{ijk} = \gamma^{1/2}\,[ijk]$.

The second complication for $K \neq 0$ is that gradients do not commute when applied to 3-vectors and 3-tensors (though they do commute for 3-scalars). The basic results are

$$\begin{aligned}
[\boldsymbol{\nabla}_j, \boldsymbol{\nabla}_k]\, A^i &= {}^{(3)}R^i{}_{njk} A^n \ , \\
[\boldsymbol{\nabla}_k, \boldsymbol{\nabla}_l]\, h^{ij} &= {}^{(3)}R^i{}_{nkl} h^{nj} + {}^{(3)}R^j{}_{nkl} h^{in} \ ,
\end{aligned} \tag{4.5}$$

where $[\boldsymbol{\nabla}_j, \boldsymbol{\nabla}_k] \equiv (\boldsymbol{\nabla}_j \boldsymbol{\nabla}_k - \boldsymbol{\nabla}_k \boldsymbol{\nabla}_j)$. The commutator involves the spatial Riemann tensor, which for a uniform-curvature space with 3-metric γ_{ij} is simply

$${}^{(3)}R^i{}_{jkl} = K\left(\delta^i{}_k \gamma_{jl} - \delta^i{}_l \gamma_{jk}\right) \ . \tag{4.6}$$

Finally, we shall need the evolution equations for the full spacetime metric $g_{\mu\nu}$. These are given by the Einstein equations,

$$G^\mu{}_\nu = 8\pi G T^\mu{}_\nu \ , \tag{4.7}$$

where $T^\mu{}_\nu$ is the stress-energy tensor and $G^\mu{}_\nu$ is the Einstein tensor, related to the spacetime Ricci tensor $R_{\mu\nu}$ by

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{R}{2} g_{\mu\nu} \ , \quad R \equiv R^\mu{}_\mu \ , \quad R_{\mu\nu} \equiv R^\kappa{}_{\mu\kappa\nu} \ . \tag{4.8}$$

The spacetime Riemann tensor is defined according to the convention

$$R^\mu{}_{\nu\kappa\lambda} \equiv \partial_\kappa \Gamma^\mu{}_{\nu\lambda} - \partial_\lambda \Gamma^\mu{}_{\nu\kappa} + \Gamma^\mu{}_{\alpha\kappa}\Gamma^\alpha{}_{\nu\lambda} - \Gamma^\mu{}_{\alpha\lambda}\Gamma^\alpha{}_{\nu\kappa} \ , \tag{4.9}$$

where the affine connection coefficients are

$$\Gamma^\mu{}_{\nu\lambda} \equiv \frac{1}{2} g^{\mu\kappa}\left(\partial_\nu g_{\kappa\lambda} + \partial_\lambda g_{\kappa\nu} - \partial_\kappa g_{\nu\lambda}\right) \ . \tag{4.10}$$

We see that the Einstein tensor involves second derivatives of the metric tensor components, so that eq. (4.7) provides second-order partial differential equations for $g_{\mu\nu}$.

The reader who is not completely comfortable with the material summarized above may wish to consult an introductory general relativity textbook, e.g. Schutz (1985).

4.2. Classification of metric perturbations

Now we consider small perturbations of the spacetime metric away from the Robertson-Walker form:

$$ds^2 = a^2(\tau)\left\{-(1+2\psi)d\tau^2 + 2w_i d\tau dx^i + \left[(1-2\phi)\gamma_{ij} + 2h_{ij}\right] dx^i dx^j\right\} ,$$
$$\gamma^{ij} h_{ij} = 0 \ . \tag{4.11}$$

We have introduced two 3-scalar fields $\psi(\boldsymbol{x},\tau)$ and $\phi(\boldsymbol{x},\tau)$, one 3-vector field $\boldsymbol{w}(\boldsymbol{x},\tau)=w_i\boldsymbol{e}^i$, and one symmetric, traceless second-rank 3-tensor field $\mathsf{h}(\boldsymbol{x},\tau)=h_{ij}\boldsymbol{e}^i\otimes\boldsymbol{e}^j$. No generality is lost by making h_{ij} traceless since any trace part can be put into ϕ. The factors of 2 and signs have been chosen to simplify later expressions.

Equation (4.11) is completely general: $g_{\mu\nu}$ has 10 independent components and we have introduced 10 independent fields ($1+1+3+5$ for $\psi+\phi+\boldsymbol{w}+\mathsf{h}$). In fact, only 6 of these fields can represent physical degrees of freedom because we are free to transform our 4 coordinates (τ,x^i) without changing any physical quantities. Infinitesimal coordinate transformations, called gauge transformations, result in changes of the fields $(\psi,\phi,\boldsymbol{w},\mathsf{h})$ because the spacetime scalar $ds^2=g_{\mu\nu}dx^\mu dx^\nu$ must be invariant under general coordinate transformations. We shall explore the consequences of this invariance later. Coordinate invariance complicates general relativity compared with other gauge theories (e.g., electromagnetism) in which the spacetime coordinates are fixed while other variables change under the appropriate gauge transformations.

Unless stated explicitly to the contrary, in the following we shall treat the perturbation variables (ψ,ϕ,w_i,h_{ij}) exclusively as 3-tensors (of rank 0, 1, or 2 according to the number of indices) with components raised and lowered using γ^{ij} and γ_{ij}. In doing this we *choose* to use γ_{ij} as the 3-metric in the perturbed hypersurface of constant τ despite the fact that the spatial part of the 4-metric (divided by a^2) is given by $(1-2\phi)\gamma_{ij}+2h_{ij}$. This treatment is satisfactory because we will assume that the metric perturbations are small and we will neglect all terms quadratic in them. However, we will use $g^{\mu\nu}$ to raise 4-vector components: $G^\mu{}_\nu=g^{\mu\kappa}G_{\kappa\nu}$. Do take care to distinguish Latin from Greek!

We have introduced 3-scalar, 3-vector, and 3-tensor perturbations. (From now on we will drop the prefix 3- since it should be clear from the context whether 3- or 4- is implied.) Are these the famous scalar, vector, and tensor metric perturbations? Not quite! Recall the decomposition of a vector into longitudinal and transverse parts:

$$\boldsymbol{w}=\boldsymbol{w}_{||}+\boldsymbol{w}_\perp\ ,\quad \boldsymbol{\nabla}\times\boldsymbol{w}_{||}=\boldsymbol{\nabla}\cdot\boldsymbol{w}_\perp=0\ . \tag{4.12}$$

Since $\boldsymbol{w}_{||}=-\boldsymbol{\nabla}w$ for some scalar w, how can it be called a vector perturbation? By definition, only the *transverse* component $\boldsymbol{w}_\perp$ represents a vector perturbation.

There is a similar decomposition theorem for tensor fields: Any differentiable traceless symmetric 3-tensor field $h_{ij}(\boldsymbol{x})$ may be decomposed into a sum of parts, called longitudinal, solenoidal, and transverse:

$$\mathsf{h}(\boldsymbol{x})=\mathsf{h}_{||}+\mathsf{h}_\perp+\mathsf{h}_{\mathrm{T}}\ . \tag{4.13}$$

The various parts are defined in terms of a scalar field $h(\boldsymbol{x})$ and transverse (or solenoidal) vector field $\boldsymbol{h}(\boldsymbol{x})$ such that

$$h_{ij,\,||} = D_{ij} h \ , \quad h_{ij,\,\perp} = \boldsymbol{\nabla}_{(i} h_{j)} \ , \quad \boldsymbol{\nabla}_i h^i{}_{j,\,\mathrm{T}} = 0 \ , \tag{4.14}$$

where we have denoted symmetrization with parentheses and have employed the traceless symmetric double gradient operator:

$$\boldsymbol{\nabla}_{(i} h_{j)} \equiv \frac{1}{2}\left(\boldsymbol{\nabla}_i h_j + \boldsymbol{\nabla}_j h_i\right) \ , \quad D_{ij} \equiv \boldsymbol{\nabla}_i \boldsymbol{\nabla}_j - \frac{1}{3}\gamma_{ij}\boldsymbol{\nabla}^2 \ . \tag{4.15}$$

Note that the divergences of $\mathsf{h}_{||}$ and $\mathsf{h}_\perp$ are longitudinal and transverse vectors, respectively (it doesn't matter which index is contracted on the divergence since h is symmetric):

$$\boldsymbol{\nabla}\cdot\mathsf{h}_{||} = \frac{2}{3}\boldsymbol{\nabla}\left(\boldsymbol{\nabla}^2 + 3K\right) h \ , \quad \boldsymbol{\nabla}\cdot\mathsf{h}_\perp = \frac{1}{2}\left(\boldsymbol{\nabla}^2 + 2K\right)\boldsymbol{h} \ , \tag{4.16}$$

where $\boldsymbol{\nabla}^2\boldsymbol{h} \equiv (\boldsymbol{\nabla}^2 h^i)\boldsymbol{e}_i$. (We do not call $\mathsf{h}_\perp$ the transverse part, as we would by extension from $\boldsymbol{w}_\perp$, because "transverse" is conventionally used to refer to the tensor part.) The longitudinal tensor $\mathsf{h}_{||}$ is also called the scalar part of h, the solenoidal part $\mathsf{h}_\perp$ is also called the vector part, and the transverse-traceless part h_T is also called the tensor part. This classification of the spatial metric perturbations h_{ij} was first performed by Lifshitz (1946).

The purpose of this decomposition is to separate h_{ij} into parts that can be obtained from scalars, vectors, and tensors. Is the decomposition unique? Not quite. It is clear, first of all, that h and h_i are defined only up to a constant. But there may be additional freedom (Stewart 1990).

First, the vector $\boldsymbol{h}$ is defined only up to solutions of Killing's equation $\boldsymbol{\nabla}_i h_j + \boldsymbol{\nabla}_j h_i = 0$, called Killing vectors (Misner et al. 1973). The reader can easily verify that one such solution (using the quasi-Cartesian coordinates of eq. 4.3) is $(h_x, h_y, h_z) = (y, -x, 0)$. In an open space ($K \leqslant 0$) this solution would be excluded because it is unbounded — our perturbations should not diverge! — but in a closed space ($K > 0$) the coordinates have a bounded range. This Killing vector, and its obvious cousins, correspond to global rotations of the spatial coordinates and not to physical perturbations.

Next, there may also be non-uniqueness associated with the tensor (and scalar) component:

$$h_{ij,\,\mathrm{T}} \to h_{ij,\,\mathrm{T}} + \zeta_{ij} \ , \quad \zeta_{ij} \equiv \left[\boldsymbol{\nabla}_i\boldsymbol{\nabla}_j - \gamma_{ij}\left(\boldsymbol{\nabla}^2 + 2K\right)\right]\zeta \ , \tag{4.17}$$

where ζ is some scalar field. From eqs. (4.5) and (4.6) one can show $\boldsymbol{\nabla}^2(\boldsymbol{\nabla}_i\zeta) = \boldsymbol{\nabla}_i(\boldsymbol{\nabla}^2 + 2K)\zeta$ so that $\boldsymbol{\nabla}_i\zeta^i{}_j = 0$ as required for the tensor component. However, we also require $h_{ij,\mathrm{T}}$ to be traceless, implying $(\boldsymbol{\nabla}^2 + 3K)\zeta = 0$. Thus, the tensor mode is defined only up to eq. (4.17) with bounded solutions of $(\boldsymbol{\nabla}^2 + 3K)\zeta = 0$. In fact, this condition also implies $\zeta_{ij} = D_{ij}\zeta$, so we may equally well attribute ζ_{ij} to the scalar mode $h_{ij,||}$. Thus, we are free to add any multiple of ζ to h (the scalar mode) provided we subtract $D_{ij}\zeta$ from the tensor mode. In an open space ($K \leqslant 0$) there are no nontrivial bounded solutions to $(\boldsymbol{\nabla}^2 + 3K)\zeta = 0$ but in a closed space ($K > 0$) there are four linearly independent solutions (Stewart 1990). Once again, these solutions correspond to redefinitions of the coordinates with no physical significance. Kodama & Sasaki (1984, Appendix B) gave a proof of the tensor decomposition theorem, but they missed the additional vector and scalar/tensor mode solutions present in a closed space. In practice, it is easy to exclude these modes, and so we shall ignore them hereafter.

Thus, we conclude that the most general perturbations of the Robertson-Walker metric may be decomposed at each point in space into four scalar parts each having 1 degree of freedom ($\psi, \phi, \boldsymbol{w}_{||}, \mathsf{h}_{||}$), two vector parts each having 2 degrees of freedom ($\boldsymbol{w}_{\perp}, \mathsf{h}_{\perp}$), and one tensor part having 2 degrees of freedom (h_{T}, which lost 3 degrees of freedom to the transversality condition). The total number of degrees of freedom is 10.

Why do we bother with this mathematical classification? First and foremost, the different metric components represent distinct physical phenomena. (By way of comparison, in previous lectures we have already seen that $\boldsymbol{v}_{||}$ and $\boldsymbol{v}_{\perp}$ play very different roles in fluid motion.) Ordinary Newtonian gravity obviously is a scalar phenomenon (the Newtonian potential is a 3-scalar), while gravitomagnetism and gravitational radiation — both of which are absent from Newton's laws, and will be discussed below — are vector and tensor phenomena, respectively. Moreover, this spatial decomposition can also be applied to the Einstein and stress-energy tensors, allowing us to see clearly (at least in some coordinate systems) the physical sources for each type of gravity. Finally, the classification will help us to eliminate unphysical gauge degrees of freedom. There are at least four of them, corresponding to two of the scalar fields and one transverse vector field.

We will not write the weak-field Einstein equations for the general metric of eq. (4.11). Instead, we will consider only two particular gauge choices, each of which allows for all physical degrees of freedom (and more, in the case of synchronous gauge). First, however, we must examine the stress-energy tensor.

4.3. Stress-energy tensor

The Einstein field eqs. (4.7) show that the stress-energy tensor provides the source for the metric variables. For a perfect fluid the stress-energy tensor takes the well-known form

$$T^{\mu\nu} = (\rho + p)u^\mu u^\nu + pg^{\mu\nu} \ , \tag{4.18}$$

where ρ and p are the proper energy density and pressure in the fluid rest frame and $u^\mu = dx^\mu/d\lambda$ (where $d\lambda^2 \equiv -ds^2$) is the fluid 4-velocity. In any locally flat coordinate system, T^{00} represents the energy density, T^{0i} the energy flux density (which equals the momentum density T^{i0}), and T^{ij} represents the spatial stress tensor. In locally flat coordinates in the fluid frame, $T^{00} = \rho$, $T^{0i} = 0$, and $T^{ij} = p\delta^{ij}$ for a perfect fluid.

For an imperfect fluid such as a sum of several uncoupled components (e.g., photons, neutrinos, baryons, and cold dark matter), the stress-energy tensor must include extra terms corresponding in a weakly collisional gas to shear and bulk viscosity, thermal conduction, and other physical processes. We may write the general form as

$$T^{\mu\nu} = (\rho + p)u^\mu u^\nu + pg^{\mu\nu} + \Sigma^{\mu\nu} \ . \tag{4.19}$$

Without loss of generality we can require $\Sigma^{\mu\nu}$ to be traceless and flow-orthogonal: $\Sigma^\mu{}_\mu = 0$, $\Sigma^\mu{}_\nu u^\nu = 0$. In locally flat coordinates in the fluid rest frame only the spatial components Σ^{ij} are nonzero (but their trace vanishes) and the spatial stress is $T^{ij} = p\delta^{ij} + \Sigma^{ij}$. With these restrictions on $\Sigma^{\mu\nu}$ (in particular, the absence of a Σ^{0i} term in the fluid rest frame) we implicitly define u^μ so that ρu^μ is the *energy current* 4-vector (as opposed, for example, to the particle mass times the *number* current 4-vector for the baryons or other conserved particles). As a result of these conditions, ρu^μ includes any heat conduction, p includes any bulk viscosity (the isotropic stress generated when an imperfect fluid is rapidly compressed or expanded), and $\Sigma^{\mu\nu}$ (called the shear stress) includes shear viscosity. Some workers add to eq. (4.19) terms proportional to the 4-velocity, $q^\mu u^\nu + u^\mu q^\nu$, where q^μ is the energy current in the particle frame (taking u^μ to be proportional to the particle number current). Either choice is fully general, although our choice is the simplest.

We shall need to evaluate the stress-energy components in the comoving coordinate frame implied by eq. (4.11). This requires specifying the form of the 4-velocity u^μ. Therefore we must digress to discuss the 4-velocity components in a perturbed spacetime.

Consider first the case where the fluid is at rest in the comoving frame, i.e., $u^i = 0$. (This condition *defines* the comoving frame.) Normalization

($g_{\mu\nu}u^\mu u^\nu = -1$) then requires $u^0 = a^{-1}(1-\psi)$ to first order in ψ. Lowering the components using the full 4-metric gives $u_0 = -a(1+\psi)$ and $u_i = aw_i$ in the weak-field approximation.

The appearance of ψ and w_i in the components u_μ for a fluid at rest in the comoving frame may appear odd. They arise because, in our coordinates, clocks run at different rates in different places if $\boldsymbol{\nabla}_i\psi \neq 0$ (the coordinate time interval $d\tau$ corresponds to a proper time interval $a(\tau)(1+\psi)d\tau$) and they also have a position-dependent offset if $w_i \neq 0$ (an observer at $x^i =$ constant sees the clocks at $x^i + dx^i$ running fast by an amount $w_i dx^i$). At first these may seem like strange coordinate artifacts one should avoid (this may be a motivation for the synchronous gauge in which $\psi = w_i = 0$!) but they have straightforward physical interpretations: ψ represents the gravitational redshift and w_i represents the dragging of inertial frames. We shall see later that they also can be interpreted as giving rise to "forces," allowing us to apply Newtonian intuition in general relativity. Do not forget that in general relativity we are forced to accept coordinates whose relation to proper times and distances is complicated by spacetime curvature. Therefore, it is advantageous when we can reinterpret these effects in Newtonian terms.

We define the coordinate 3-velocity

$$\boldsymbol{v} \equiv \frac{d\boldsymbol{x}}{d\tau} = \frac{dx^i}{dx^0}\,\boldsymbol{e}_i = \frac{u^i}{u^0}\,\boldsymbol{e}_i \ , \tag{4.20}$$

whose components are to be raised and lowered using γ^{ij} and γ_{ij}: $v_i = \gamma_{ij}v^j = \gamma_{ij}u^j/u^0$, $v^2 \equiv \gamma_{ij}v^iv^j$, $\boldsymbol{w}\cdot\boldsymbol{v} \equiv w_iv^i$, $\boldsymbol{v}\cdot\mathsf{h}\cdot\boldsymbol{v} \equiv h_{ij}v^iv^j$, etc. The 4-vector component u^0 follows from applying the normalization condition $u_\mu u^\mu = -1$:

$$u^0 = \frac{1}{a\sqrt{1-v^2}}\left[1 - \frac{\psi - \boldsymbol{w}\cdot\boldsymbol{v} + \phi v^2 - \boldsymbol{v}\cdot\mathsf{h}\cdot\boldsymbol{v}}{1-v^2}\right] \ . \tag{4.21}$$

In the absence of metric perturbations this looks like the standard result in special relativity aside from the factor a^{-1} that appears because we use comoving coordinates. With metric perturbations we can no longer interpret $\boldsymbol{v}$ exactly as the *proper* 3-velocity because adx^i is not proper distance and $ad\tau$ is not proper time. However, the corrections are only first order in the metric perturbations.

We will assume that the mean fluid velocity is nonrelativistic so that we can neglect all terms that are quadratic in $\boldsymbol{v}$. (This does not exclude the radiation era, since we allow individual particles to be relativistic and require only the bulk velocity to be nonrelativistic.) We will also neglect terms involving products of $\boldsymbol{v}$ and the metric perturbations. With these approximations, the 4-velocity components become

$$u^0 = a^{-1}(1-\psi) \ , \ \ u^i = a^{-1}v^i \ , \ \ u_0 = -a(1+\psi) \ , \ \ u_i = a(v_i+w_i) \ . \tag{4.22}$$

The apparent lack of symmetry in the spatial components arises because $u_i = g_{i0}u^0 + g_{ij}u^j$ and $g_{i0} = a^2 w_i \neq 0$ in general.

From eq. (4.22) we can see how w_i is interpreted as a frame-dragging effect. For $w_i \neq 0$ the worldline of a comoving observer (defined by the condition $v_i = 0$) is not normal to the hypersurfaces $\tau =$ constant: $u_\mu \xi^\mu = a w_i \xi^i \neq 0$ for a 3-vector ξ^i. In a locally inertial frame, on the other hand, the worldline of a freely-falling observer obviously would be normal to the spatial directions. (This is true in special relativity and also in general relativity as a consequence of the equivalence principle.) By making a local Galilean transformation, $dx^i \to dx^i + w^i d\tau$, we can remove w_i from the metric at a point. This transformation corresponds to choosing a locally inertial frame, called the *normal* frame, moving with 3-velocity $-\boldsymbol{w}$ relative to the comoving frame. In the normal frame the fluid 3-velocity is $\boldsymbol{v} + \boldsymbol{w}$.

If $w_i = w_i(\tau)$ is independent of $\boldsymbol{x}$, one can remove w_i everywhere from the metric by a global Galilean transformation. (Try it and see!) However, we may be interested in situations where $w_i = w_i(\boldsymbol{x}, \tau)$ so that different transformations are required in different places. In this case there is no *global* inertial frame. Spatially varying w_i corresponds to shearing and/or rotation of the comoving frame relative to the normal frame. This is called the "dragging of inertial frames." Although we can choose coordinates in which $w_i = 0$ everywhere, we shall see that there are advantages in not hiding the dragging of inertial frames. In general, the comoving frame is noninertial: an observer can remain at fixed x^i only if accelerated by nongravitational forces. The synchronous gauge is an exception in that $w_i = 0$ everywhere and the comoving frame is locally inertial. We shall see later that these features of synchronous gauge obscure rather than eliminate the physical dragging of inertial frames.

Now that we have all the ingredients we can finally write the stress-energy tensor components in our perturbed comoving coordinate system in terms of physical quantities:

$$\begin{gathered} T^0{}_0 = -\rho \ , \ \ T^i{}_0 = -(\rho + p)v^i \ , \\ T^0{}_i = (\rho + p)(v_i + w_i) \ \ , \ \ T^i{}_j = p\delta^i{}_j + \Sigma^i{}_j \ . \end{gathered} \tag{4.23}$$

We use mixed components in order to avoid extraneous factors of $a(1+\psi)$ and $a(1-\phi)$. Note that the traceless shear stress $\Sigma^i{}_j$ may be decomposed as in eqs. (4.13) and (4.14) into scalar, vector, and tensor parts. Similarly, the energy flux density $(\rho+p)v^i$ may be decomposed into scalar and vector parts. (The pressure appears here, just as in special relativity, to account for the pdV work done in compressing a fluid. For a nonrelativistic fluid $p \ll \rho$, but we shall not make this restriction.) We may already anticipate that these sources are responsible in the Einstein equations for scalar, vector, and tensor metric perturbations.

In writing the components of the stress-energy tensor we have not assumed $|\delta\rho| \ll \bar{\rho}$. The only approximations we make in the stress-energy tensor are to neglect (relative to unity) v^2 and all terms involving products of the metric perturbations with $\boldsymbol{v}$ and $\Sigma^i{}_j$. Of course, owing to the weak-field approximation, we are also neglecting any terms that are quadratic in the metric perturbations themselves.

Before moving on to discuss the Einstein equations we should rewrite the conservation of energy-momentum, $\boldsymbol{\nabla}_\mu T^\mu{}_\nu = 0$, in terms of our metric perturbation and fluid variables. (We use $\boldsymbol{\nabla}_\mu$ to denote the full spacetime covariant derivative relative to the 4-metric $g_{\mu\nu}$. It should not be confused with the spatial gradient $\boldsymbol{\nabla}_i$ defined relative to the 3-metric γ_{ij}.) Using the approximations mentioned in the preceding paragraph, one finds

$$\partial_\tau \rho + 3(\eta - \dot{\phi})(\rho + p) + \boldsymbol{\nabla} \cdot [(\rho + p)\boldsymbol{v}\,] = 0 \tag{4.24}$$

and

$$\begin{aligned}\partial_\tau \left[(\rho + p)(\boldsymbol{v} + \boldsymbol{w})\right] \; &+ 4\eta(\rho + p)(\boldsymbol{v} + \boldsymbol{w}) \\ &+ \boldsymbol{\nabla} p + \boldsymbol{\nabla} \cdot \Sigma + (\rho + p)\boldsymbol{\nabla}\psi = 0 \; .\end{aligned} \tag{4.25}$$

(Deriving these gives useful practice in tensor algebra.) It is easy to interpret the various terms in these equations. The terms proportional to the expansion rate η arise because we are using comoving coordinates and conformal time and have not factored out a^{-3} from ρ or p. The pressure p is present with ρ because we let ρ be the energy density (not the rest-mass density), which is affected by the work pressure does in compressing the fluid. Excluding these terms, the energy-conservation eq. (4.24) looks exactly like the Newtonian continuity equation aside from the change in the expansion rate from η to $\eta - \dot{\phi}$. This modification is easily understood by noting from eq. (4.11) that the effective isotropic expansion factor is modified by spatial curvature perturbations to become $a(1 - \phi)$. The momentum-conservation eq. (4.25) similarly looks like the Newtonian version with a gravitational potential ψ, aside from the special-relativistic effects of pressure and the addition of $\boldsymbol{w}$ to all the velocities to place them in the normal (inertial) frame.

4.4. Synchronous gauge

Synchronous gauge, introduced by Lifshitz (1946) in his pioneering calculations of cosmological perturbation theory, is defined by the conditions $\psi = w_i = 0$, which eliminate two scalar fields (ψ and the longitudinal part of $\boldsymbol{w}$) and one transverse vector field ($\boldsymbol{w}_\perp$). It is not difficult to show that synchronous coordinates can be found for any weakly-perturbed spacetime. However, the synchronous gauge conditions do not eliminate all gauge freedom. This has in the past led to considerable confusion (for discussion see Press & Vishniac 1980 and Bardeen 1980).

Synchronous gauge has the property that there exists a set of comoving observers who fall freely without changing their spatial coordinates. (This is nontrivial when one notes that in order to remain at a fixed terrestrial latitude, longitude, and altitude above the surface of the earth it is necessary to accelerate everywhere except in geosynchronous orbits.) These observers are called "fundamental" comoving observers. The existence of fundamental observers follows from the geodesic equation

$$\frac{du^\mu}{d\lambda} + \Gamma^\mu{}_{\alpha\beta} u^\alpha u^\beta = 0 \tag{4.26}$$

for the trajectory $x^\mu(\lambda)$, where $d\lambda = (-ds^2)^{1/2}$ for a timelike geodesic and $u^\mu = dx^\mu/d\lambda$. With $\psi = w_i = 0$, eq. (4.10) gives $\Gamma^i{}_{00} = 0$, implying that $u^i = 0$ is a geodesic.

Each fundamental observer carries a clock reading conformal time $\tau = \int dt/a(t)$ and a fixed spatial coordinate label x^i. The clocks and labels of the fundamental observers are taken to *define* the coordinate values at all spacetime points (assuming that these hypothetical observers densely fill space). The residual gauge freedom in synchronous gauge arises from the freedom to adjust the initial settings of the clocks and the initial coordinate labels of the fundamental observers.

Because the spatial coordinates x^i of each fundamental observer are held fixed with time, the x^i in synchronous gauge are Lagrangian coordinates. This implies that the coordinate lines become highly deformed when the density perturbations become large. When the trajectories of two fundamental observers intersect the coordinates become singular: two different sets of x^μ label the same spacetime event. This flaw of synchronous gauge is not apparent if $|\delta\rho/\bar\rho| \ll 1$ and the initial coordinate labels are nearly unperturbed, so this gauge may be used successfully (with some care required to avoid contamination of physical variables by the residual gauge freedom) in linear perturbation theory.

To be consistent with the conventional notation used for synchronous gauge (Lifshitz 1946; Lifshitz & Khalatnikov 1963; Weinberg 1972; Peebles 1993), in this section only we shall absorb ϕ into h_{ij} and double h_{ij}:

$$ds^2 = a^2(\tau)\left[-d\tau^2 + (\gamma_{ij} + h_{ij})dx^i dx^j\right] \ , \quad h \equiv h^i{}_i \neq 0 \ . \tag{4.27}$$

Using this line element and the definitions of the Ricci and Einstein tensors, it is straightforward (if rather tedious) to derive the components of the perturbed Einstein tensor:

$$-a^2 G^0{}_0 = 3(\eta^2 + K) + \eta\dot{h} - \frac{1}{2}\left(\boldsymbol{\nabla}^2 + 2K\right)h + \frac{1}{2}\boldsymbol{\nabla}_i\boldsymbol{\nabla}_j h^{ij} \ , \tag{4.28}$$

$$a^2 G^0{}_i = \frac{1}{2}\left(\boldsymbol{\nabla}_i\dot{h} - \boldsymbol{\nabla}_j\dot{h}^j{}_i\right) \ , \quad G^i{}_0 = -\gamma^{ij}G^0{}_j \ , \tag{4.29}$$

$$\begin{aligned} -a^2 G^i{}_j = \ & \left(2\dot{\eta} + \eta^2 + K\right)\delta^i{}_j + \left(\frac{1}{2}\partial_\tau^2 + \eta\partial_\tau - \frac{1}{2}\boldsymbol{\nabla}^2\right)\left(h\delta^i{}_j - h^i{}_j\right) \\ & -Kh^i{}_j + \frac{1}{2}\gamma^{ik}\left(\boldsymbol{\nabla}_k\boldsymbol{\nabla}_j h - \boldsymbol{\nabla}_k\boldsymbol{\nabla}_l h^l{}_j - \boldsymbol{\nabla}_j\boldsymbol{\nabla}_l h^l{}_k\right) \\ & +\frac{1}{2}\left(\boldsymbol{\nabla}_k\boldsymbol{\nabla}_l h^{kl}\right)\delta^i{}_j \ . \end{aligned} \tag{4.30}$$

One can easily verify that the unperturbed parts of the Einstein equations $G^0{}_0 = 8\pi G T^0{}_0 = -8\pi G\bar{\rho}$ and $G^i{}_j = 8\pi G T^i{}_j = 8\pi G\bar{p}\delta^i{}_j$ give the Friedmann and energy-conservation equations for the background Robertson-Walker spacetime.

Our next goal is to separate the perturbed Einstein equations into scalar, vector, and tensor parts. First we must decompose the metric perturbation field h_{ij} as in eqs. (4.13)–(4.15), with a term added (and the notation changed slightly) to account for the trace of h_{ij}:

$$h_{ij} = \frac{1}{3}h\gamma_{ij} + D_{ij}\left(\boldsymbol{\nabla}^{-2}\xi\right) + \boldsymbol{\nabla}_{(i}h_{j)} + h_{ij,\mathrm{T}} \ , \tag{4.31}$$

where D_{ij} was defined in eq. (4.15). We require $\boldsymbol{\nabla}_i h^i = \boldsymbol{\nabla}_i h^i{}_{j,\mathrm{T}} = 0$ to ensure that the last two parts of h_{ij} are purely solenoidal (vector mode) and transverse-traceless (tensor mode) contributions. The scalar mode variables are h and $\boldsymbol{\nabla}^{-2}\xi$, whose Laplacian is ξ. We shall not worry about how to invert the Laplacian on a curved space but simply assume that it can be done if necessary.

The perturbed Einstein equations now separate into 7 different parts according to the spatial symmetry:

$$G^0{}_0: \quad \frac{1}{3}\left(\boldsymbol{\nabla}^2 + 3K\right)(\xi - h) + \eta\dot{h} = 8\pi G a^2(\rho - \bar{\rho}) \,, \tag{4.32}$$

$$G^0{}_{i,\,\|}: \quad \frac{1}{3}\boldsymbol{\nabla}_i(\dot{h} - \dot{\xi}) - K\boldsymbol{\nabla}_i\left(\boldsymbol{\nabla}^{-2}\dot{\xi}\right) = 8\pi G a^2\left[(\rho + p)v_i\right]_{\|} \,, \tag{4.33}$$

$$G^0{}_{i,\,\perp}: \quad -\frac{1}{4}\left(\boldsymbol{\nabla}^2 + 2K\right)\dot{h}_i = 8\pi G a^2\left[(\rho + p)v_i\right]_{\perp} \,, \tag{4.34}$$

$$G^i{}_i: \quad -(\partial_\tau^2 + 2\eta\partial_\tau)h + \frac{1}{3}\left(\boldsymbol{\nabla}^2 + 3K\right)(h - \xi) = 24\pi G a^2(p - \bar{p}) \,, \tag{4.35}$$

$$G^i{}_{j\neq i,\,\|}: \quad \left(\frac{1}{2}\partial_\tau^2 + \eta\partial_\tau\right)D_{ij}\left(\boldsymbol{\nabla}^{-2}\xi\right) + \frac{1}{6}D_{ij}(\xi - h) = 8\pi G a^2\Sigma_{ij,\,\|} \,, \tag{4.36}$$

$$G^i{}_{j,\,\perp}: \quad \left(\frac{1}{2}\partial_\tau^2 + \eta\partial_\tau\right)\boldsymbol{\nabla}_{(i}h_{j)} = 8\pi G a^2\Sigma_{ij,\,\perp} \,, \tag{4.37}$$

$$G^i{}_{j,\,\mathrm{T}}: \quad \left(\frac{1}{2}\partial_\tau^2 + \eta\partial_\tau - \frac{1}{2}\boldsymbol{\nabla}^2 + K\right)h_{ij,\,\mathrm{T}} = 8\pi G a^2\Sigma_{ij,\,\mathrm{T}} \,. \tag{4.38}$$

The derivation of these equations is straightforward but tedious. They have decomposed naturally into separate equations for the scalar, vector, and tensor parts of the metric perturbation, with the sources for each given by the appropriate part of the energy-momentum tensor. However, there are more equations than unknowns! There are four scalar equations for ξ and h, two vector equations for h_i, and one tensor equation for $h_{ij,\,\mathrm{T}}$. How can this be?

Before answering this question, let us note another interesting feature of the equations above, which will provide a clue. The equations arising from $G^0{}_\mu$ involve only a single time derivative of the scalar and vector mode variables, while those arising from $G^i{}_\mu$ have two time derivatives, as we might have expected for equations of motion for the gravitational fields. This means that we could discard eqs. (4.32)–(4.34) and be left with exactly as many second-order in time equations as unknown fields. Alternatively, we could discard eqs. (4.35)–(4.37) and be left with exactly enough first-order in time equations for the scalar and vector modes. Only the tensor mode evolution is uniquely specified by a second-order wave equation.

The reason for this redundancy is that the twice-contracted Bianchi identities of differential geometry, $\boldsymbol{\nabla}_\mu G^\mu{}_\nu = 0$, force the Einstein eqs. (4.7) to imply $\nabla_\mu T^\mu{}_\nu = 0$. The Einstein equations themselves contain redundancy, as we can check explicitly here. By combining the time derivative of eq. (4.32) and the divergence of eqs. (4.33) and (4.34) one obtains the perturbed part of eq. (4.24) (note, however, that $\phi \to -h/6$). Similarly, eq. (4.25) follows from the time derivative of eqs. (4.33) and (4.34) combined with the gradient of eqs. (4.35)–(4.37). *Because we require the equations of motion for the matter and radiation to locally conserve the net energy-momentum, three of the perturbed Einstein eqs.* (4.32)–(4.38) *are redundant.*

In the literature, $G^0{}_0 = 8\pi G T^0{}_0$ is often called the "ADM energy constraint" and $G^0{}_i = 8\pi G T^0{}_i$ is called the "ADM momentum constraint" equation. The 3+1 space-time decomposition of the Einstein equations into constraint and evolution equations was developed in detail by Arnowitt, Deser & Misner (1962, ADM) and applied to cosmology by Durrer & Straumann (1988) and Bardeen (1989). The ADM constraint equations may be regarded as providing initial-value constraints on $(h, \xi, \dot{h}, \dot{\xi}, \dot{h}_i)$ and the matter variables. If these constraints are satisfied initially (this is required for a consistent metric), and if eqs. (4.35)–(4.37) are used to evolve $(h, \xi, \dot{h}, \dot{\xi}, \dot{h}_i)$ while the matter variables are evolved so as to locally conserve the net energy-momentum, then the ADM constraints will be fulfilled at all later times. (This follows from the results stated in the preceding paragraph.) In effect, the Einstein equations have built into themselves the *requirement* of energy-momentum conservation for the matter. If one were to integrate eqs. (4.35)–(4.37) correctly but to violate energy-momentum conservation, then eqs. (4.32)–(4.34) would be violated.

In practice, we may find it preferable to regard the ADM constraints alone — and not eqs. (4.35)–(4.37) — as giving the actual field equations for the scalar and vector metric perturbations. They have fewer time derivatives and hence are easier to integrate. Equations (4.35)–(4.37) are not necessary at all (although they may be useful for numerical checks) because they can always be obtained by differentiating eqs. (4.32)–(4.34) and using energy-momentum conservation.

This situation becomes clearer if we compare it with Newtonian gravity. The field equation $\nabla^2\phi = 4\pi G a^2 \delta\rho$ is analogous to eq. (4.32). (We shall see this equivalence much more clearly in the Poisson gauge below.) Let us take the time derivative: $\nabla^2\dot{\phi} = 4\pi G \partial_\tau(a^2\delta\rho)$. If we now replace $\partial_\tau(\delta\rho)$ using the continuity equation, we obtain a time evolution equation for $\nabla^2\dot{\phi}$ analogous to the divergence of eq. (4.33). The solutions to this evolution equation obey the Poisson equation if and only if the initial ϕ obeys the Poisson equation. Why should one bother to integrate $\nabla^2\dot{\phi}$ in time when the solution can always be obtained instantaneously from the

Poisson equation? Viewed in this way, we would say that the extra time derivatives in the $G^i{}_\mu$ equations have nothing to do with gravity *per se*. The *real* field equations for the scalar and vector modes come from the ADM constraint equations.

If the scalar and vector metric perturbations evolve according to first-order in time equations, their solutions are not manifestly causal (e.g., retarded solutions of the wave equation). We shall discuss this point in detail in section 4.7. However, for now we may note that the tensor mode obeys the wave eq. (4.38). The solutions are the well-known gravity waves which, as we shall see, play a key role in enforcing causality. The source for these waves is given by the transverse-traceless stress (generated, for example, by two masses orbiting around each other). The $\eta\partial_\tau$ term arises because we use comoving coordinates and the K term arises as a correction to the Laplacian in a curved space; otherwise the vacuum solutions are clearly waves propagating at the speed of light. Abbott & Harari (1986) show that eq. (4.38) is the Klein-Gordon equation for a massless spin-two particle.

4.5. Gauge modes

As we noted above, the synchronous gauge conditions do not completely fix the spacetime coordinates because of the freedom to redefine the perturbed constant-time hypersurfaces and to reassign the spatial coordinates within these hypersurfaces. This freedom is not obvious in the linearized Einstein equations for the scalar and vector modes, but it is present in the form of additional solutions that must be fixed by appropriate choice of initial conditions and that represent nothing more than relabeling of the coordinates in an unperturbed Robertson-Walker spacetime.

To see this effect more clearly, we consider a general infinitesimal coordinate transformation from (τ, x^i) to $(\hat{\tau}, \hat{x}^i)$, known as a **gauge transformation**:

$$\hat{\tau} = \tau + \alpha(\boldsymbol{x}, \tau) \ , \quad \hat{x}^i = x^i + \gamma^{ij}\boldsymbol{\nabla}_j\beta(\boldsymbol{x}, \tau) + \epsilon^i(\boldsymbol{x}, \tau) \ , \quad \text{with} \ \ \boldsymbol{\nabla}\cdot\epsilon = 0 \ . \tag{4.39}$$

For convenience we have split the spatial transformation into longitudinal and transverse parts. Note that the transformed time and space coordinates depend in general on all four of the old coordinates.

Coordinate freedom leads to ambiguity in the meaning of density perturbations. Consider, for example, the simple case of an unperturbed Robertson-Walker universe in which the density depends only on τ (if one uses the "correct" τ coordinate). In the transformed system it depends also on $\hat{x}^i$: $\bar{\rho}(\hat{\tau}) = \bar{\rho}(\tau) + (\partial_\tau\bar{\rho})\alpha(\boldsymbol{x},\tau)$. In other words, even in an unperturbed universe we can be fooled into thinking there are spatially-varying density perturbations.

This example may seem contrived, but the ambiguity is not trivial to avoid: When spacetime itself is perturbed, and time is not absolute, what is the best choice of time? The same question arises for the spatial coordinates.

To clarify this situation we must examine gauge transformations further. First note that when we transform the coordinates we must also transform the metric perturbation variables so that the line element ds^2 (a spacetime scalar) is invariant. It is straightforward to do this using eqs. (4.11) and (4.39). The result is

$$\hat{\psi} = \psi - \dot{\alpha} - \eta\alpha \ , \quad \hat{\phi} = \phi + \frac{1}{3}\boldsymbol{\nabla}^2\beta + \eta\alpha \ ,$$
$$\hat{w}_i = w_i + \boldsymbol{\nabla}_i(\alpha - \dot{\beta}) - \dot{\epsilon}_i \ , \quad \hat{h}_{ij} = h_{ij} - D_{ij}\beta - \boldsymbol{\nabla}_{(i}\epsilon_{j)} \ , \tag{4.40}$$

where D_{ij} is the traceless double gradient operator defined in eq. (4.15). The transformed fields (with carets) are to be evaluated at the same coordinate values (τ, x^i) as the original fields.

Suppose now that our original coordinates satisfy the synchronous gauge conditions $\psi = w_i = 0$. [To recover the notation of eq. (4.27) used specially for synchronous gauge we now double h_{ij} and put the trace of h_{ij} into $h = -6\phi$.] From eqs. (4.40) and (4.27) it follows that there is a whole *family* of synchronous gauges with metric variables related to the original ones by

$$\hat{h} = h - 2\boldsymbol{\nabla}^2\beta - 6\eta\dot{\beta} \ , \quad \hat{\xi} = \xi - 2\boldsymbol{\nabla}^2\beta \ , \quad \hat{h}_i = h_i - 2\epsilon_i \ , \tag{4.41}$$

where

$$\beta = \beta_0(\boldsymbol{x})\int\frac{d\tau}{a(\tau)} \ , \quad \epsilon_i = \epsilon_i(\boldsymbol{x}) \ . \tag{4.42}$$

Thus, the *synchronous gauge has residual freedom in the form of one scalar (β_0) and one transverse vector (ϵ_i) function of the spatial coordinates.*

The presence of these extraneous solutions (called gauge modes) has created a great deal of confusion in the past, which might have been avoided had more cosmologists read the paper of Lifshitz (1946). In 1980, Bardeen wrote an influential paper showing how one may take linear combinations of the metric and matter perturbation variables that are free of gauge modes. For example, Bardeen defined two scalar perturbations Φ_A and Φ_H related to our synchronous gauge variables h and ξ (Bardeen actually used the variables $H_L \equiv h/6$ and $H_T \equiv -\xi/2$) as follows:

$$\Phi_A = -\frac{1}{2}\,\boldsymbol{\nabla}^{-2}(\ddot{\xi} + \eta\dot{\xi}) \ , \quad \Phi_H = \frac{1}{6}\,(h - \xi) - \frac{1}{2}\,\eta\boldsymbol{\nabla}^{-2}\dot{\xi} \ . \tag{4.43}$$

It is easy to check that these variables are invariant under the synchronous gauge transformation given by eqs. (4.41)–(4.42).

Bardeen's work led to a flurry of papers concerning gauge-invariant variables in cosmology. A standard reference is the classic paper by Kodama & Sasaki (1984). Elegant treatments based on general 3+1 splitting of spacetime were given later by Durrer & Straumann (1988) and Bardeen (1989). The simpler form of the gauge-invariant variables often makes it easier to find analytical solutions (e.g., Rebhan 1992). However, it is not *necessary* to use gauge-invariant variables during a calculation, and many cosmologists have continued successfully to use synchronous gauge. In the end, when the results are converted to measurable quantities — spacetime scalars — the gauge modes automatically get canceled. In a numerical solution, however, one must be careful that the gauge modes do not swamp the physical ones, otherwise roundoff can produce significant numerical errors.

Gauge invariant variables actually appear somewhat strange if we consider the analogous situation in electromagnetism. The electric and magnetic fields in flat spacetime may be obtained from potentials ϕ and $\boldsymbol{A}$ (note we are implicitly using a 3+1 split of spacetime),

$$\boldsymbol{E} = -\boldsymbol{\nabla}\phi - \partial_\tau \boldsymbol{A} \ , \quad \boldsymbol{B} = \boldsymbol{\nabla} \times \boldsymbol{A} \ . \tag{4.44}$$

With this choice, the source-free Maxwell equations are automatically satisfied; the other two (the Coulomb and Ampère laws) become

$$\boldsymbol{\nabla}^2\phi + \partial_\tau(\boldsymbol{\nabla}\cdot\boldsymbol{A}) = -4\pi\rho \ , \quad \left(\partial_\tau^2 - \boldsymbol{\nabla}^2\right)\boldsymbol{A} + \boldsymbol{\nabla}\dot{\phi} = 4\pi\boldsymbol{J} \ , \tag{4.45}$$

where ρ is the charge density and $\boldsymbol{J}$ is the current density. These equations are invariant under the gauge transformation $\hat{\phi} = \phi - \partial_\tau\alpha$, $\hat{A}_i = A_i + \boldsymbol{\nabla}_i\alpha$.

If we didn't know about electric and magnetic fields, but were alarmed by the gauge-dependence of the potentials, we could try to find linear combinations of ϕ and $\boldsymbol{A}$ that are gauge-invariant. However, there are two well-known and more direct ways to eliminate gauge modes. The first is "gauge fixing" — i.e., placing constraints on the potentials so as to eliminate gauge degrees of freedom. One popular choice, for example, is the Coulomb gauge $\boldsymbol{\nabla}\cdot\boldsymbol{A} = 0$, so that $\boldsymbol{A} = \boldsymbol{A}_\perp$ is transverse. The transversality condition means that the gauge transformation variable α cannot depend on position (though it can depend on time); thus, most of the gauge freedom is eliminated. The second possibility is to work with the physical fields themselves instead of the potentials: $\boldsymbol{E}$ and $\boldsymbol{B}$ are automatically gauge-invariant. This procedure requires that we analyze the equation of motion for charges to determine which combinations of ϕ and $\boldsymbol{A}$ are physically most significant.

In the next section we shall adopt the first procedure (gauge-fixing) using the gravitational analogue of the Coulomb gauge. Later we shall introduce Ellis' covariant approach based on gravitational fields themselves.

4.6. Poisson gauge

Recall that our general perturbed Robertson-Walker metric (4.11) contains four extraneous degrees of freedom associated with coordinate invariance. In the synchronous gauge these degrees of freedom are eliminated from g_{00} (one scalar) and g_{0i} (one scalar and one transverse vector) by requiring $\psi = w_i = 0$. There are other ways to eliminate the same number of fields. As we shall see, a good choice is to constrain g_{0i} (eliminating one scalar) and g_{ij} (eliminating one scalar and one transverse vector) by imposing the following gauge conditions on eq. (4.11):

$$\boldsymbol{\nabla} \cdot \boldsymbol{w} = 0 \ , \quad \boldsymbol{\nabla} \cdot \mathsf{h} = 0 \ . \tag{4.46}$$

I call this choice the **Poisson gauge** by analogy with the Coulomb gauge of electromagnetism ($\boldsymbol{\nabla} \cdot \boldsymbol{A} = 0$).† More conditions are required here than in electromagnetism because gravity is a *tensor* rather than a *vector* gauge theory. Note that in the Poisson gauge there are *two scalar* potentials (ψ and ϕ), *one transverse vector* potential ($\boldsymbol{w}$), and *one transverse-traceless* tensor potential h.

A restricted version of the Poisson gauge, with $w_i = h_{ij} = 0$, is known in the literature as the longitudinal or conformal Newtonian gauge (Mukhanov, Feldman & Brandenberger 1992). These conditions can be applied only if the stress-energy tensor contains no vector or tensor parts and there are no free gravitational waves, so that only the scalar metric perturbations are present. While this condition may apply, in principle, in the linear regime ($|\delta\rho/\bar{\rho}| \ll 1$), nonlinear density fluctuations generally induce vector and tensor modes even if none were present initially. Setting $\boldsymbol{w} = \mathsf{h} = 0$ is analogous to zeroing the electromagnetic vector potential,

† The same gauge has been proposed recently by Bombelli, Couch & Torrence (1994), who call it "cosmological gauge." However, I prefer the name Poisson gauge because cosmology — i.e., nonzero $\dot{a}$ — is irrelevant for the definition and physical interpretation of this gauge. Although I have seen no earlier discussion of Poisson gauge in the literature, its time slicing corresponds with the minimal shear hypersurface condition of Bardeen (1980).

implying $\boldsymbol{B} = 0$. In general, this is not a valid *gauge condition* — it is rather the elimination of physical phenomena. The longitudinal/conformal Newtonian gauge really should be called a "restricted gauge." The Poisson gauge, by contrast, allows all physical degrees of freedom present in the metric.

To prove the last statement, and to find out how much residual gauge freedom is allowed, we must find a coordinate transformation from an *arbitrary* gauge to the Poisson gauge. Using eq. (4.40) with hats indicating Poisson gauge variables, we see that a suitable transformation exists with

$$\alpha = w + \dot{h} \ , \quad \beta = h \ , \quad \epsilon_i = h_i \ , \tag{4.47}$$

where w comes from the longitudinal part of $\boldsymbol{w}$ ($\boldsymbol{w}_{\|} = -\boldsymbol{\nabla} w$), while h and h_i come from the longitudinal and solenoidal parts of h in eq. (4.14). Because these conditions are *algebraic* in α, β, and ϵ (they are not differentiated, in contrast with the transformation to synchronous gauge of eq. 4.41), we have found an almost unique transformation from an arbitrary gauge to the Poisson gauge. One can still add arbitrary functions of time alone (with no dependence on x^i) to α and ϵ_i. (Adding a function of time alone to β has no effect at all because the transformation, eq. 4.39, involves only the gradient of β.)

Spatially homogeneous changes in α represent changes in the units of time and length, while spatially homogeneous changes in ϵ represent shifts in the origin of the spatial coordinate system. These trivial residual gauge freedoms — akin to electromagnetic gauge transformations generated by a function of time, the only gauge freedom remaining in Coulomb gauge — are physically transparent and should cause no conceptual or practical difficulty.

It is interesting to see the coordinate transformation from a synchronous gauge to the Poisson gauge. As an exercise the reader can show that this is given by

$$\psi = -\frac{1}{2}\,\boldsymbol{\nabla}^{-2}(\ddot{\xi}+\eta\dot{\xi}) \ , \quad \phi = \frac{1}{6}\,(\xi-h)+\frac{1}{2}\,\eta\boldsymbol{\nabla}^{-2}\dot{\xi} \ , \quad w_i = -\frac{1}{2}\,\partial_\tau h_i \ . \tag{4.48}$$

Comparing with eq. (4.43), we see that the two Poisson-gauge scalar potentials are $\psi = \Phi_A$ and $\phi = -\Phi_H$. (Kodama & Sasaki 1984 call these variables $\Psi = \psi$ and $\Phi = -\phi$.) The vector potential w_i in Poisson gauge is related simply to the solenoidal potential h_i of the synchronous gauge (eq. 4.31).

Thus, the metric perturbations in the Poisson gauge correspond exactly with several of the gauge-invariant variables introduced by Bardeen. By imposing the explicit gauge conditions (4.46), we have simplified the mathematical analysis of these variables.

Now that we have seen that the Poisson gauge solves the gauge-fixing problem, let us give the components of the perturbed Einstein equations. They are no more complicated than those of the synchronous gauge:

$$G^0{}_0: \quad \left(\boldsymbol{\nabla}^2 + 3K\right)\phi - 3\eta\left(\dot{\phi} + \eta\psi\right) = 4\pi G a^2(\rho - \bar{\rho}) \ , \tag{4.49}$$

$$G^0{}_{i,\,\|}: \quad -\boldsymbol{\nabla}_i(\dot{\phi} + \eta\psi) = 4\pi G a^2 \left[(\rho + p)(v_i + w_i)\right]_{\|} \ , \tag{4.50}$$

$$G^0{}_{i,\,\perp}: \quad \left(\boldsymbol{\nabla}^2 + 2K\right) w_i = 16\pi G a^2 \left[(\rho + p)(v_i + w_i)\right]_{\perp} \ , \tag{4.51}$$

$$\begin{aligned} G^i{}_i: \quad & \ddot{\phi} - K\phi + \eta(\dot{\psi} + 2\dot{\phi}) + (2\dot{\eta} + \eta^2)\psi - \frac{1}{3}\boldsymbol{\nabla}^2(\phi - \psi) \\ & = 4\pi G a^2(p - \bar{p}) \ , \end{aligned} \tag{4.52}$$

$$G^i{}_{j\neq i,\,\|}: \quad D_{ij}\left(\phi - \psi\right) = 8\pi G a^2 \Sigma_{ij,\,\|} \ , \tag{4.53}$$

$$G^i{}_{j,\,\perp}: \quad -\left(\partial_\tau + 2\eta\right)\boldsymbol{\nabla}_{(i} w_{j)} = 8\pi G a^2 \Sigma_{ij,\,\perp} \ , \tag{4.54}$$

$$G^i{}_{j,\,\mathrm{T}}: \quad \left(\partial_\tau^2 + 2\eta\partial_\tau - \boldsymbol{\nabla}^2 + 2K\right) h_{ij} = 8\pi G a^2 \Sigma_{ij,\,\mathrm{T}} \ . \tag{4.55}$$

As in the synchronous gauge, the scalar and vector modes satisfy initial-value (ADM) constraints (eqs. 4.49–4.51) in addition to evolution equations. However, it is remarkable that in the Poisson gauge we can obtain the scalar and vector potentials *directly* from the instantaneous stress-energy distribution with no time integration required. This is clear for $\phi - \psi$ and $\boldsymbol{w}$, both of which obey elliptic equations with no time derivatives (eqs. 4.53 and 4.51, respectively). By combining the ADM energy and longitudinal momentum constraint equations we can also get an instantaneous equation for ϕ:

$$\left(\boldsymbol{\nabla}^2 + 3K\right)\phi = 4\pi G a^2 \left[\delta\rho + 3\eta\Phi_f\right] \ , \quad -\boldsymbol{\nabla}\Phi_f \equiv \left[(\rho + p)(\boldsymbol{v} + \boldsymbol{w})\right]_{\|} \ . \tag{4.56}$$

Bardeen (1980) defined the matter perturbation variable $\epsilon_m \equiv (\delta\rho + 3\eta\Phi_f)/\bar{\rho}$ and noted that it is the natural measure of the energy density fluctuation in the normal (inertial) frame at rest with the matter such that $\boldsymbol{v} + \boldsymbol{w} = 0$ (recall the discussion in section 4.3). However, for our analysis we will remain in the comoving frame of the Poisson gauge, in which case $\delta\rho/\bar{\rho}$ and not ϵ_m is the density fluctuation.

We can show that for nonrelativistic matter the field equations we have obtained reduce to the Newtonian forms. First, it is clear that in the non-cosmological limit ($\eta = K = 0$), eq. (4.56) reduces to the Poisson equation. For $\eta \neq 0$ the longitudinal momentum density Φ_f is also a source for ϕ, but it is unimportant for perturbations with $|\delta\rho/\bar{\rho}| \gg v_H v/c^2$ where v_H is the

Hubble velocity across the perturbation. Next, consider the implications of the fact that the shear stress for any physical system is *at most* $O(\rho c_{\rm s}^2)$ where $c_{\rm s}$ is the characteristic thermal speed of the gas particles. (For a collisional gas the shear stress is much less than this.) Equation (4.53) then implies that the relative difference between ψ and ϕ is no more than $O(c_{\rm s}/c)^2$. Third, eq. (4.51) implies that the vector potential $\boldsymbol{w} \sim (v_H/c)^2\boldsymbol{v}$. Thus, the deviations from the Newtonian results are all $O(v/c)^2$. *Poisson gauge gives the relativistic cosmological generalization of Newtonian gravity.*

There are still more remarkable features of the Poisson gauge. First, the Poisson gauge metric perturbation variables are almost always small in the nonrelativistic limit ($|\phi| \ll c^2$, $v^2 \ll c^2$), in contrast with the synchronous gauge variables h_{ij}, which become large when $|\delta\rho/\bar{\rho}| > 1$. (However, Bardeen 1980 shows that the relative numerical merits of these two gauges can reverse for isocurvature perturbations of size larger than the Hubble distance.) Second, if $(\psi, \phi, \boldsymbol{w}, \mathsf{h})$ are very small, they — but not necessarily their derivatives! — may be neglected to a good approximation, in which case the Poisson gauge coordinates reduce precisely to the Eulerian coordinates used in Newtonian cosmology. Finally, it is amazing that the scalar and vector potentials depend solely on the *instantaneous* distribution of stress-energy — in fact, only the energy and momentum densities and the shear stress are required. *Only the tensor mode* — gravitational radiation — *follows unambiguously from a time evolution equation.* In fact, it obeys precisely the same equation as in the synchronous gauge (with a factor of 2 difference owing to our different definitions) because tensor perturbations are gauge-invariant — coordinate transformations involving 3-scalars and a 3-vector cannot change a 3-tensor (leaving aside the special case of eq. 4.17 for a closed space).

4.7. Physical content of the Einstein equations

In the last section we showed that the Poisson gauge variables $(\psi, \phi, \boldsymbol{w})$ are given by the *instantaneous* distributions of energy density, momentum density, and shear stress (longitudinal momentum flux density). Is this action at a distance in general relativity?

We showed in eq. (4.47) that the Poisson gauge can be transformed to any other gauge. In the cosmological Lorentz gauge (see Misner et al. 1973 for the noncosmological version) all metric perturbation components obey wave equations. Therefore, the solutions in Poisson gauge *must be causal* despite appearances to the contrary.

There is a precedent for this type of behavior: the Coulomb gauge of electromagnetism. With $\boldsymbol{\nabla}\cdot\boldsymbol{A}=0$, eqs. (4.45) become

$$\boldsymbol{\nabla}^2\phi=-4\pi\rho\ ,\quad \boldsymbol{\nabla}\dot{\phi}=4\pi\boldsymbol{J}_{||}\ ,\quad \left(\partial_\tau^2-\boldsymbol{\nabla}^2\right)\boldsymbol{A}=4\pi\boldsymbol{J}_\perp\ . \tag{4.57}$$

We have separated the current density into longitudinal and transverse parts. The similarity of the first two (scalar) equations to eqs. (4.49) and (4.50) is striking. The similarity would be even more striking if we were to use comoving coordinates rather than treating $\boldsymbol{x}$ and τ here as flat spacetime coordinates. As an exercise one can show that with comoving coordinates, ρ and $\boldsymbol{J}$ will be multiplied by a^2 and that $\dot{\phi}$ becomes $\dot{\phi}+\eta\phi$. The last step follows when one distinguishes time derivatives at fixed $\boldsymbol{x}$ from those at fixed $a\boldsymbol{x}$.

Are we to conclude that electromagnetism *also* violates causality, because the electric potential ϕ depends only on the instantaneous distribution of charge? No! To understand this let us examine the Coulomb and Ampère laws in flat spacetime for the *fields* rather than the potentials:

$$\boldsymbol{\nabla}\cdot\boldsymbol{E}=\boldsymbol{\nabla}\cdot\boldsymbol{E}_{||}=4\pi\rho\ ,\ -\partial_\tau\boldsymbol{E}_{||}=4\pi\boldsymbol{J}_{||}\ ,\ \boldsymbol{\nabla}\times\boldsymbol{B}-\partial_\tau\boldsymbol{E}_\perp=4\pi\boldsymbol{J}_\perp\ . \tag{4.58}$$

The Ampère law has been split into longitudinal and transverse parts. We see that the *longitudinal* electric field indeed *is* given instantaneously by the charge density. Because the photon is a massless vector particle, only the *transverse* part of the electric and magnetic fields is radiative, and its source is given by the transverse current density:

$$\left(\partial_\tau^2-\boldsymbol{\nabla}^2\right)\boldsymbol{B}=4\pi\boldsymbol{\nabla}\times\boldsymbol{J}_\perp\ ,\quad \left(\partial_\tau^2-\boldsymbol{\nabla}^2\right)\boldsymbol{E}_\perp=-4\pi\partial_\tau\boldsymbol{J}_\perp\ . \tag{4.59}$$

But how does this restore causality? To see how, let us consider the following example. Suppose that there is only one electric charge in the universe and initially it is at rest in the lab frame. If the charge moves — even much more slowly than the speed of light — $\boldsymbol{E}_{||}$ — the solution to the Coulomb equation — is changed everywhere instantaneously. It must be therefore that $\boldsymbol{E}_\perp$ *also* changes instantaneously in such a way as to exactly *cancel* the acausal behavior of $\boldsymbol{E}_{||}$.

This indeed happens, as follows. First, note that the motion of the charge generates a current density $\boldsymbol{J}=\boldsymbol{J}_{||}+\boldsymbol{J}_\perp$. The longitudinal and transverse parts separately extend over all space (and are in this sense acausal) while their sum vanishes away from the charge (as do $\boldsymbol{\nabla}\cdot\boldsymbol{J}_{||}$ and $\boldsymbol{\nabla}\times\boldsymbol{J}_\perp$). The magnetic and transverse electric fields obey eqs. (4.59). Because $\boldsymbol{J}_\perp$ is distributed over all space but $\boldsymbol{\nabla}\times\boldsymbol{J}_\perp$ is not, retarded-wave solutions for $\boldsymbol{B}$ are localized and causal while those for $\boldsymbol{E}_\perp$ are not. However, when $\boldsymbol{E}_{||}$ is added to $\boldsymbol{E}_\perp$, one finds that the net electric field is causal (Brill & Goodman 1967). It is a useful exercise to show this in detail.

Now that we understand how causality is maintained, what is the use of the longitudinal part of the Ampère law, $-\partial_\tau \boldsymbol{E}_{||} = 4\pi \boldsymbol{J}_{||}$? The answer is, to ensure charge conservation, which is implied by combining the time derivative of the Coulomb law with the divergence of the Ampère law:

$$\partial_\tau \rho + \boldsymbol{\nabla} \cdot \boldsymbol{J} = \partial_\tau \rho + \boldsymbol{\nabla} \cdot \boldsymbol{J}_{||} = 0 \ . \tag{4.60}$$

Charge conservation is built into the Coulomb and Ampère laws. This remarkable behavior occurs because electromagnetism is a *gauge* theory. Gauge invariance effectively provides a redundant scalar field equation whose physical role is to enforce charge conservation. From Noether's theorem (e.g., Goldstein 1980), a continuous symmetry (in this case, electromagnetic gauge invariance) leads to a conserved current.

General relativity is also a gauge theory. Coordinate invariance — a continuous symmetry — leads to conservation of energy and momentum. As a result there are redundant scalar and vector equations [eqs. (4.50), (4.52), and (4.54)] *whose role is to enforce the conservation laws* [eqs. (4.24) and (4.25)]. We are free to use the action-at-a-distance field equations for the scalar and vector potentials in Poisson gauge because, when they are converted to fields and combined with the gravitational radiation field, the resulting behavior is entirely causal.

The analogy with electromagnetism becomes clearer if we replace the gravitational potentials by fields. We define the "gravitoelectric" and "gravitomagnetic" fields (Thorne, Price & Macdonald 1986; Jantzen, Carini & Bini 1992)

$$\boldsymbol{g} = -\boldsymbol{\nabla}\psi - \partial_\tau \boldsymbol{w} \ , \quad \boldsymbol{H} = \boldsymbol{\nabla} \times \boldsymbol{w} \ , \tag{4.61}$$

using the Poisson gauge variables ψ and $\boldsymbol{w}$. In section 4.8 we shall see how these fields lead to "forces" on particles similar to the Lorentz forces of electromagnetism. For now, however, we are interested in the fields themselves.

Note that $\boldsymbol{g}$ and $\boldsymbol{H}$ are invariant under the transformation $\psi \to \psi - \dot{\alpha}$, $\boldsymbol{w} \to \boldsymbol{w} + \boldsymbol{\nabla}\alpha$. In the noncosmological limit ($\eta = 0$) this is a gauge transformation corresponding to transformation of the time coordinate (cf. eqs. 4.39 and 4.40). However, gauge transformations in general relativity are complicated by the fact that they change the coordinates and fields as well as the potentials. For example, the $\eta\alpha$ terms in eq. (4.40) arise because the transformed metric is evaluated at the old coordinates. Thus, $\boldsymbol{g}$ should acquire a term $\eta\boldsymbol{\nabla}\alpha$ under a true gauge (coordinate) transformation, which is incompatible with eq. (4.61). The actual transformation ($\psi \to \psi - \dot{\alpha}$, $\boldsymbol{w} \to \boldsymbol{w} + \boldsymbol{\nabla}\alpha$) is *not* a coordinate transformation. General relativity differs from electromagnetism in that gauge transformations change not just the potentials but also the coordinates used to evaluate the potentials; remember that the potentials *define* the perturbed coordinates! Only in a simple coordinate system, such as Poisson gauge — the gravitational analogue of Coulomb gauge — is it possible to see a simple relation between fields and potentials similar to that of electromagnetism.

In the limit of comoving distance scales small compared with the curvature distance $|K|^{-1/2}$ and the Hubble distance η^{-1}, and nonrelativistic shear stresses, the gravitoelectric and gravitomagnetic fields obey a gravitational analogue of the Maxwell equations:

$$\begin{aligned}&\boldsymbol{\nabla}\cdot\boldsymbol{g}=-4\pi Ga^2\delta\rho\ ,\quad \boldsymbol{\nabla}\times\boldsymbol{g}+\partial_\tau\boldsymbol{H}=0\ ,\\ &\boldsymbol{\nabla}\cdot\boldsymbol{H}=0\ ,\quad \boldsymbol{\nabla}\times\boldsymbol{H}=-16\pi Ga^2\boldsymbol{f}_\perp\ ,\end{aligned}\tag{4.62}$$

where $\boldsymbol{f}=(\rho+p)(\boldsymbol{v}+\boldsymbol{w})$ is the momentum density in the normal (inertial) frame. (You may derive these equations using eqs. 4.49, 4.50, 4.53, and 4.61.) These equations differ from their electromagnetic counterparts in three essential ways: (1) the sources have opposite sign (gravity is attractive), (2) the transverse momentum density has a coefficient 4 times larger than the transverse electric current (gravity is a tensor and not a vector theory), and (3) there is no "displacement current" $-\partial_\tau\boldsymbol{g}$ in the transverse Ampère law for $\boldsymbol{\nabla}\times\boldsymbol{H}$. Recalling that Maxwell added the electric displacement current precisely to conserve charge and thereby obtained radiative (electromagnetic wave) solutions, we understand the difference here: the vector component of gravity is nonradiative. Unlike the photon, the graviton is a spin-2 particle (or would be if we could quantize general relativity!), so radiative solutions appear only for the (transverse-traceless) tensor potential h_{ij}. In fact, the vector potential is nonradiative precisely because it is needed to ensure momentum conservation; mass conservation is already taken care of by the scalar potential. Recall the role of the ADM constraint equations discussed in section 4.4. Gravity has more conservation laws to maintain than electromagnetism and consequently needs more fields to constrain.

Obtaining this physical insight into general relativity is *much* easier in the Poisson gauge than in the synchronous gauge. This fact alone is a good reason for preferring the former. When combined with the other advantages (simpler equations, no time evolution required for the scalar and vector potentials, reduction to the Newtonian limit, no nontrivial gauge modes, and lack of unphysical coordinate singularities), the superiority of the Poisson gauge should be clear.

Although the physical picture we have developed for gravity in analogy with electromagnetism is beautiful, it is inexact. Not only have we linearized the metric, we have also neglected cosmological effects in eqs. (4.62). We shall see in section 4.9 how to obtain exact nonlinear equations for (the gradients of) the gravitational fields.

4.8. Hamiltonian dynamics of particles

In this section we extend to general relativity the Hamiltonian formulation of particle dynamics that is familiar in Newtonian mechanics. In the process we shall obtain further insight into the physical meaning of the gravitational fields discussed in the previous section. A preliminary version of this material appears in (Bertschinger 1993). A related presentation in the context of gravitational fields near black holes is given by Thorne et al. (1986).

As in the nonrelativistic case, we choose a Hamiltonian that is related to the energy of a particle. Consequently, our approach is not manifestly covariant; the energy depends on how spacetime is sliced into hypersurfaces of constant conformal time τ because the energy is the time component of a 4-vector. Nevertheless, our approach is fully compatible with general relativity; we must only select a specific gauge. For simplicity we shall adopt the Poisson gauge, eq. (4.11) with gauge conditions eq. (4.46). We assume that the metric perturbations are given by a solution of the field eqs. (4.49)–(4.55). Our Hamiltonian will include only the degrees of freedom associated with one particle; one can generalize this to include many particles (even treated as a continuum) and the metric variables (Arnowitt et al. 1962; Misner et al. 1973; Salopek & Stewart 1992) but this involves more machinery than necessary for our purposes.

The goal of the Hamiltonian approach is to obtain equations of motion for trajectories in the single-particle phase space consisting of the spatial coordinates x^i and their conjugate momenta. The first question is, what are the appropriate conjugate momenta? This question practically answers itself when we express the action scalar in terms of our coordinates:

$$S = \int P_\mu dx^\mu = \int \left(P_\tau + P_i \frac{dx^i}{d\tau} \right) d\tau = \int (-H + P_i \dot{x}^i)\, d\tau \ . \qquad (4.63)$$

Note that we have automatically expressed the action in terms of the covariant (lower-index) components of the 4-momentum (also known as the components of the momentum one-form). We can read off the Hamiltonian and conjugate momenta using the fact that $S = \int L d\tau$ where $L(x^i, \dot{x}^j, \tau)$ is the Lagrangian, which is related to the Hamiltonian $H(x^i, P_j, \tau)$ by the Legendre transformation $L = P_i \dot{x}^i - H$. The Hamiltonian therefore is $H = -P_\tau$ — despite appearances, we shall see that this is *not* in general the *proper* energy — and the conjugate momenta equal the covariant spatial components of the 4-momentum. Indeed, we may simply *define* the conjugate momenta and Hamiltonian in this way. (Care should be taken not to confuse the Hamiltonian H with the Hubble parameter H and the gravitomagnetic field $\boldsymbol{H}$!)

With these definitions, H and P_i correspond to the usual quantities encountered in elementary nonrelativistic mechanics, but we need not rely on this fact. For any choice of spacetime geometry and coordinates we may determine the corresponding Hamiltonian and conjugate momenta from the 4-momentum components: for a particle of mass m, $H = -mg_{0\mu}dx^\mu/d\lambda$, $P_i = mg_{i\mu}dx^\mu/d\lambda$ where $d\lambda$ measures proper time along the particle trajectory. As an exercise, one may show that with cylindrical coordinates (r, θ, z) for a nonrelativistic particle of mass m in Minkowski spacetime, $P_r = m\dot{r}$ is the radial momentum, $P_\theta = mr^2\dot{\theta}$ is the angular momentum about $\boldsymbol{e}_z$, $P_z = m\dot{z}$ is the linear momentum along $\boldsymbol{e}_z$, and $H = E \approx m + (P_r^2 + P_\theta^2/r^2 + P_z^2)/2m$ is the proper energy (including the rest mass energy). We shall determine the functional form $H(x^i, P_j, \tau)$ for our perturbed Robertson-Walker spacetime below.

First, however, let us show that our approach leads to the usual canonical Hamilton's equations of motion, rigorously justifying our choices $H = -P_\tau$ and P_i being the momentum conjugate to x^i. To do this we simply vary the phase space trajectory $\{x^i(\tau),\, P_j(\tau)\}$ to $\{x^i + \delta x^i,\, P_j + \delta P_j\}$, treating $\delta x^i(\tau)$ and $\delta P_j(\tau)$ as independent variations and computing the variation of the action of eq. (4.63):

$$\begin{aligned}
\delta S &= \int \left(-\frac{\partial H}{\partial x^i}\delta x^i - \frac{\partial H}{\partial P_i}\delta P_i + \frac{dx^i}{d\tau}\delta P_i + P_i \frac{d}{d\tau}\delta x^i \right) d\tau \\
&= \int \left[\left(\frac{dx^i}{d\tau} - \frac{\partial H}{\partial P_i} \right) \delta P_i(\tau) - \left(\frac{dP_i}{d\tau} + \frac{\partial H}{\partial x^i} \right) \delta x^i(\tau) \right] d\tau \ , \qquad (4.64)
\end{aligned}$$

where we have assumed $P_i \delta x^i = 0$ at the endpoints of integration. Requiring the action to be stationary under all variations, $\delta S = 0$, we obtain the standard form of Hamilton's equations:

$$\frac{dx^i}{d\tau} = \frac{\partial H}{\partial P_i} \ , \quad \frac{dP_i}{d\tau} = -\frac{\partial H}{\partial x^i} \ . \qquad (4.65)$$

Thus, Hamilton's equations give phase space trajectories in general relativity just as they do in nonrelativistic mechanics.

Our next step is to determine the Hamiltonian for the problem at hand. We shall assume that the particle falls freely in the perturbed Robertson-Walker spacetime described in the Poisson gauge. For comparison with the nonrelativistic results, it is useful to relate the 4-momentum components to the proper energy and 3-momentum measured by a comoving observer (i.e., one at fixed x^i), E and p_i:

$$P_\tau = -a(1+\psi)E \ , \quad P_i = a\left[(1-\phi)(p_i + Ew_i) + h_{ij}p^j\right] \ . \qquad (4.66)$$

The first equation follows from $E = -u^\mu P_\mu$ where u^μ is the 4-velocity of a comoving observer from eq. (4.21) with $\boldsymbol{v} = 0$, while the second equation follows from projecting P_μ into the hypersurface normal to u^μ and normalizing to give the proper 3-momentum. The weak-field approximation has been made (i.e., terms quadratic in the metric perturbations are neglected), but the particle motion is allowed to be relativistic. The factors $a(1+\psi)$ and $a(1-\phi)$ are obviously needed from eq. (4.11) to convert proper quantities into coordinate momenta, the Ew_i term arises because our space and time coordinates are not orthogonal if there is a vector mode, and the $h_{ij}p^j$ term arises because our spatial coordinates are not orthogonal if there is a tensor mode. The reader may verify that the 4-momentum satisfies the normalization condition $g^{\mu\nu}P_\mu P_\nu = -E^2 + p^2 = -m^2$, and that this condition would be violated in general without the vector and tensor terms in P_i.

Using these results it is easy to show that, to first order in the metric perturbations, the Hamiltonian is

$$H(x^i, P_j, \tau) = \left[|(1+\phi)\boldsymbol{P} - \epsilon\,\boldsymbol{w} - \mathsf{h}\cdot\boldsymbol{P}|^2 + a^2m^2\right]^{1/2} + \epsilon\,\psi \ , \qquad (4.67)$$

where

$$\epsilon = \epsilon(\boldsymbol{P}, \tau) \equiv \left(P^2 + a^2m^2\right)^{1/2} \qquad (4.68)$$

and the squares and dot products of 3-vectors such as p_i, P_i, and $h^{ij}P_j$ are computed using the 3-metric, e.g., $P^2 = \gamma^{ij}P_iP_j$. Using the Hamiltonian of eq. (4.67), eqs. (4.65) may be shown to be fully equivalent to the geodesic equations for a freely falling particle moving in the metric of eq. (4.11), and they could also be obtained starting from a Lagrangian approach. The advantage of the Hamiltonian approach is that it treats positions and conjugate momenta equally as is needed for a phase space description.

Equation (4.67) appears strange at first glance. To understand it better, let us recall the standard form for the Hamiltonian of a particle with charge e in electromagnetic fields (with ϕ being the electrostatic potential):

$$H_e(x^i, P_j, t) = \left[|\boldsymbol{P} - e\boldsymbol{A}|^2 + m^2\right]^{1/2} + e\phi \ . \qquad (4.69)$$

Note that the proper momentum is $\boldsymbol{p} = \boldsymbol{P} - e\boldsymbol{A}$ where $\boldsymbol{P}$ is the conjugate momentum. Comparing eqs. (4.67) and (4.69), we see that they are very similar aside from the tensor term $\mathsf{h} \cdot \boldsymbol{P}$ present in the gravitational case. The few remaining differences are easily understood. To compensate for spatial curvature — effectively a local change of the units of length — in the gravitational case $\boldsymbol{P}$ is multiplied by $(1 + \phi)$. The electric charge e is replaced by the gravitational charge ϵ (energy!); to zeroth order in the perturbations $\epsilon = H = aE$. The use of comoving coordinates is responsible for the factors of $a(\tau)$. The gravitational (gravitomagnetic) vector potential is $\boldsymbol{w}$ — as we anticipated in eq. (4.61). Finally, the electrostatic potential energy $e\phi$ is replaced in the gravitational case by $\epsilon\,\psi$. The strong analogy between the vector mode and magnetism accounts for the adjective "gravitomagnetic."

A different interpretation of the gravitomagnetic contribution to the Hamiltonian will clarify the relation of gravitomagnetism and the dragging of inertial frames. In section 4.3 we noted that $\boldsymbol{w}$ is the velocity of the comoving frame relative to a locally inertial frame (the normal frame). For $w^2 \ll 1$, $\boldsymbol{p}' \equiv \boldsymbol{p} + E\boldsymbol{w}$ is therefore the proper momentum in the normal frame. According to eq. (4.66), then, neglecting the scalar and tensor modes, $\boldsymbol{P}$ is the comoving momentum (i.e., multiplied by a) in the *normal* frame, $\boldsymbol{P} = a\boldsymbol{p}'$, while $\boldsymbol{P} - \epsilon\,\boldsymbol{w}$ (the combination present in the Hamiltonian) is the comoving momentum in the *comoving* frame. It is logical that the Hamiltonian should depend on the latter quantity; after all, we are using non-orthogonal comoving spacetime coordinates. However, it is equally reasonable that the conjugate momentum should be measured in the frame normal to the hypersurface τ = constant. Thus, it is simply the offset between these two frames — if one likes, the dragging of inertial frames — that is responsible for the $-\epsilon\,\boldsymbol{w}$ term in eq. (4.67). Gravitomagnetism — and similarly magnetism, if one interprets $(e/m)\boldsymbol{A}$ as a velocity — can be viewed as a kinematical effect!

The tensor mode, corresponding to gravitational radiation, gives an extra term in the Hamiltonian — really in the relation between the proper and conjugate momenta — that is not present in the case of electromagnetism. Geometrically, h corresponds simply to a local volume-preserving deformation of the spatial coordinate lines, and in this way it simply extends the effect of the spatial curvature term $\phi\boldsymbol{P}$ in eq. (4.67) (ϕ represents an orientation-preserving dilatation of the coordinate lines). However, what is more important is the dynamical effect of these terms, neither of which is familiar in either Newtonian gravity or electromagnetism.

To study the dynamics of particle motion we use Hamilton's eqs. (4.65) with the Hamiltonian of eq. (4.67). In terms of the proper momentum $\boldsymbol{p}$ measured by a comoving observer, Hamilton's equations in the Poisson gauge become

$$\frac{d\boldsymbol{x}}{d\tau} = (1+\psi+\phi-\mathsf{h}\cdot)\,\frac{\boldsymbol{p}}{E'}\ , \quad E' \equiv \left[|\boldsymbol{p}+E\boldsymbol{w}|^2+m^2\right]^{1/2}\ ,$$
$$\frac{d}{d\tau}\left[a\,(1-\phi+\mathsf{h}\cdot)\,\boldsymbol{p}\right] = \epsilon\left[\boldsymbol{g}+\boldsymbol{v}\times\boldsymbol{H}-v^2\,\boldsymbol{\nabla}\phi+v^iv^j\,\boldsymbol{\nabla}h_{ij})\right]-\dot{\epsilon}\,\boldsymbol{w}\ , \tag{4.70}$$

where we have defined E' to be the proper energy in the normal frame, $\boldsymbol{v}$ is the peculiar velocity (in the weak-field limit it doesn't matter whether it is the coordinate or proper peculiar velocity nor whether it is measured in the comoving or normal frame) and $\boldsymbol{g}$ and $\boldsymbol{H}$ are the gravitoelectric and gravitomagnetic fields given by eqs. (4.61). The dot following h indicates the three-dimensional dot product, with $\mathsf{h}\cdot\boldsymbol{p}$ being a 3-vector.

Equations (4.70) appear rather complicated at first but each term can be understood without much difficulty. First, note that the factor $(1+\psi+\phi-\mathsf{h}\cdot)$ in the first equation is present solely to convert from a proper velocity to a coordinate velocity $d\boldsymbol{x}/d\tau$ according to the metric eq. (4.11). Using the transformation from the normal (primed) to comoving frame, $\boldsymbol{p}=\boldsymbol{p}'-E\boldsymbol{w}\approx\boldsymbol{p}'-E'\boldsymbol{w}$, the equation for $d\boldsymbol{x}/d\tau$ implies that the proper velocity in the comoving frame must equal $\boldsymbol{p}/E'=\boldsymbol{p}'/E'-\boldsymbol{w}$. This is identically true because $\boldsymbol{p}'/E'$ is the proper velocity in the normal frame, whose velocity relative to the comoving frame is $-\boldsymbol{w}$.

Similarly, the factor $a(1-\phi+\mathsf{h}\cdot)$ in the momentum equation simply converts the proper momentum $\boldsymbol{p}$ to the comoving momentum in the comoving frame, $\boldsymbol{P}-\epsilon\,\boldsymbol{w}$ (cf. eq. 4.66). The first two terms on the right-hand side have exactly the same form as the Lorentz force law of electrodynamics, with the electric charge e replaced by the comoving energy ϵ and the electric and magnetic fields $\boldsymbol{E}$ and $\boldsymbol{B}$ replaced by their gravitational counterparts $\boldsymbol{g}$ and $\boldsymbol{H}$. Thus, general relativity in the weak-field limit gives "forces" on freely-falling bodies (when expressed in the Poisson gauge) that are very similar to those of electromagnetism!

The remaining terms in the momentum equation have no counterpart in electrodynamics or Newtonian gravity. There are two gravitational force terms quadratic in the velocity arising from spatial curvature. The first one is present for a scalar mode and is responsible for the fact that photons are deflected twice as much as nonrelativistic particles in a gravitostatic field ($\phi=\psi$ in the Newtonian limit). The second term represents, in effect, scattering of moving particles by gravitational radiation. A gravity wave traveling in the z-direction will accelerate a particle in this direction if the particle has nonzero velocity in the x-y plane (the direction of polarization of the transverse gravity wave). If the particle is at rest in our coordinate system, it remains at rest when a gravity wave passes by. However, because the gravity wave corresponds to a deformation of the spatial coordinate lines, the *proper* distance between two particles at rest in the coordinate system does change (Misner et al. 1973).

Finally, the last term in the momentum equation, $-\dot{\epsilon}\,\boldsymbol{w}$, represents a sort of cosmic drag that causes velocities of massive particles to tend toward zero in the normal (inertial) frame (by driving $\boldsymbol{p}$ toward $-E\boldsymbol{w}$). The timescale for this term (the time over which ϵ changes appreciably) is the Hubble time, so it should not be regarded as the frame dragging normally spoken of when loosely describing the vector mode. In fact, in the normal frame this term is absent, but then the gravitomagnetic term changes from $\epsilon\boldsymbol{v}\times\boldsymbol{H}$ to $\epsilon\boldsymbol{\nabla}(\boldsymbol{w}\cdot\boldsymbol{v})$. The relative velocity of the comoving and normal (inertial) frames $\boldsymbol{w}$ is responsible for the frame-dragging and other effects; let us consider a particularly interesting one.

In general, $\boldsymbol{w}$ varies with position so that at different places the inertial frames rotate relative to the comoving frame with angular velocity $-\frac{1}{2}\boldsymbol{\nabla}\times\boldsymbol{w} = -\frac{1}{2}\boldsymbol{H}$; this is easily shown from a first-order Taylor series expansion of $\boldsymbol{w}$ with the constraint $\boldsymbol{\nabla}\cdot\boldsymbol{w} = 0$. As a result, a spin $\boldsymbol{S}$ will precess relative to the comoving frame at a rate $d\boldsymbol{S}/d\tau = -\frac{1}{2}\boldsymbol{H}\times\boldsymbol{S}$ (the Lense-Thirring effect). Using the magnetic analogy, one would predict a gravitomagnetic precession rate $\gamma\boldsymbol{S}\times\boldsymbol{H}$ in the comoving frame, where γ is the gyrogravitomagnetic ratio. (The analogous magnetic precession rate is $\mu\times\boldsymbol{B}$, where $\mu = \gamma\boldsymbol{S}$.) Note that this result leads to the conclusion that there is a universal gyrogravitomagnetic ratio $\gamma = \frac{1}{2}$!

Thus, one may interpret the vector mode perturbation variable $\boldsymbol{w}$ either as a source for (rather mysterious) frame-dragging effects, or as a vector potential for the gravitomagnetic field $\boldsymbol{H}$. In the former case one can eliminate $\boldsymbol{w}$ altogether by choosing orthogonal space and time coordinates such as given by the synchronous gauge. However, I prefer the latter interpretation because of the close analogy it brings to electrodynamics, allowing us to transfer our flat spacetime intuition to general relativity. The price to pay is that one must be careful to distinguish the comoving and normal frames.

We have discussed the gravitomagnetic and gravitational wave contributions to the equations of motion in order to illustrate the similarities and differences between gravity and electrodynamics. (They are clearest in the Poisson gauge; the interested reader may wish to rederive the results of this section in synchronous or some other gauge.) Why aren't we familiar with these forces in the Newtonian limit? The answer is because the sources of $\boldsymbol{H}$ and h are smaller than the source of the "gravitostatic" field $-\boldsymbol{\nabla}\psi$

by $O(v/c)$ and $O(v/c)^2$, respectively (cf. eqs. 4.62 and 4.55). From eqs. (4.70), the forces they induce are smaller by additional factors of $O(v/c)$ and $O(v/c)^2$. Thus, for nonrelativistic sources and particles, the dynamical effects of gravitomagnetism and gravitational radiation are negligible. While ordinary magnetic effects are suppressed by the same powers of v/c, the existence of opposite electric charges leads in most cases to a nearly complete cancellation of the electric charge density but not the current density. No such cancellation occurs with gravity because energy density is always positive.

Since typical gravitational fields in the universe have $\psi \approx \phi \sim 10^{-5}$ and h_{ij} is much smaller than this, the curvature factors $(1 + \psi + \phi - \mathsf{h})$ and $(1 - \phi + \mathsf{h})$ may be replaced by unity to high precision in eqs. (4.70) (and they are absent anyway in locally flat comoving coordinates). In the weak-field and slow-motion limit, then, eqs. (4.70) reduce to the standard Newtonian equations of motion in comoving coordinates.

4.9. Lagrangian field equations

General relativity makes no fundamental distinction between time and space, although we do. To obtain field equations that are similar to those of Newtonian gravity and electrodynamics, we have until now employed a "3+1 split" of the Einstein and energy conservation equations. Ellis (1971, 1973), following earlier work of Ehlers (1961, 1971), Kundt & Trümper (1961), and Hawking (1966), has developed an alternative approach based on a "1+3 split" of the Bianchi and Ricci identities. The cosmological applications have been developed extensively by Ellis and others in recent years (Ellis & Bruni 1989; Hwang & Vishniac 1990; Lyth & Stewart 1990; Bruni, Dunsby & Ellis 1992; and references therein). Ellis' approach has some important advantages, as we shall see.

The 3+1 split corresponds to the "slicing" of spacetime into a series of spatial hypersurfaces, each labeled by a coordinate time τ. (The different splitting procedures are most easily visualized with one spatial dimension suppressed using a 2+1 spacetime diagram, with time corresponding to the vertical axis. The spatial hypersurfaces are then horizontal slices through spacetime.) Spacetime is described by Eulerian observers sitting in these hypersurfaces with constant spatial coordinates.

The 1+3 split, called "threading," is complementary to slicing (Jantzen et al. 1992). In this case the fundamental geometrical objects used for charting spacetime are a series of timelike worldlines $x^\mu(\lambda; \boldsymbol{q})$, where λ is an affine parameter measuring proper time along the worldline and $\boldsymbol{q}$ gives a unique label (e.g., a spatial Lagrangian position vector) to each different worldline (or "thread"). In this case spacetime is described by Lagrangian observers moving along these worldlines.

The threading description is more general than the slicing one. If we take the threads to correspond to the worldlines of comoving observers in the slicing framework (lines of fixed $\boldsymbol{x}$), then the two descriptions are the same. In the 1+3 description, however, different threads may cross with no harmful consequences while in the 3+1 description a spatial hypersurface must not be allowed to cross itself or other slices. Thus, the threading description may be used to follow the evolution of cold dust beyond the time when matter trajectories intersect, when the perfect-fluid Euler equations break down. The advantage of a Lagrangian description is well known for collisionless matter — the Lagrangian approach exclusively is used for nonlinear gravitational simulations — and the same advantages accrue even when describing the spacetime geometry itself.

In the 1+3 approach each worldline threading spacetime has a timelike unit tangent vector (4-velocity) $u^\mu = dx^\mu/d\lambda = u^\mu(\lambda;\boldsymbol{q})$ such that $u^\mu u_\mu = -1$. Spacetime tensors are then decomposed into parts parallel and normal to the worldline passing through a given point. This decomposition is accomplished in a covariant form using the tangent vector u^μ and the orthogonal projection tensor

$$P_{\mu\nu}(u) = g_{\mu\nu} + u_\mu u_\nu \ , \tag{4.71}$$

such that $P_{\mu\nu}u^\nu = 0$ and $P^{\mu\kappa}P_{\kappa\nu} = P^\mu{}_\nu$. $P_{\mu\nu}$ is effectively the spatial metric for observers moving with 4-velocity u^μ (Ellis 1973). We may use it and u^μ to split any 4-vector A^μ into timelike and spacelike parts, labeled by the tangent vector of the appropriate thread:

$$A(u) \equiv -u_\mu A^\mu \ , \quad A^\mu(u) \equiv P^\mu{}_\nu A^\nu \ . \tag{4.72}$$

Even though $A^\mu(u)$ looks like (and is, in fact) a 4-vector, we can regard it as a 3-vector in the rest frame of an observer moving along the worldline $x^\mu(\lambda;\boldsymbol{q})$ because $u_\mu A^\mu(u) = 0$. [Note that A^μ denotes the original 4-vector while $A^\mu(u)$ denotes its projection normal to u^μ. We shall include the argument (u) for the projection whenever needed to remove ambiguity.] We require that at each point in spacetime there is at least one thread with corresponding tangent $u^\mu(\lambda;\boldsymbol{q})$. If there are several threads then there are several different decompositions of $A(u)$ and $A^\mu(u)$ at x^μ, each labeled by $\boldsymbol{q}$ (implicitly, if not explicitly) through $u^\mu(\lambda;\boldsymbol{q})$. This causes no problems as long as we refer to a single distinct thread, which we do by retaining u in the argument list.

The decomposition of a second-rank tensor $T^{\mu\nu}$ is similar:

$$T(u) = u_\mu u_\nu T^{\mu\nu} \ , \quad T_\mu(u) = g_{\mu\nu} T^\nu(u) = -P_{\mu\nu} u_\alpha T^{\alpha\nu} \ ,$$
$$T^\mu{}_\nu(u) = P^\mu{}_\alpha P_{\nu\beta} T^{\alpha\beta} \ . \tag{4.73}$$

As an exercise one may apply this decomposition to the stress-energy tensor of eq. (4.19) using the comoving observers to define the threading. For $v^2 \ll 1$, one obtains nonzero elements $T = \rho$, $T_i = a(\rho + p)v_i$ (with no w_i), and $T^i{}_j = p\delta^i{}_j + \Sigma^i{}_j$. Be careful to distinguish the 4-velocity of the threads (with $v^i = 0$) from those of the matter (eq. 4.22).

Now that we have described the 1+3 spacetime splitting procedure, we are ready to apply it to gravity following Hawking (1966) and Ellis (1971, 1973). What equations should we use? One might think to split the Einstein equations using 1+3 threading, but this does not add anything fundamentally new to what we have already done. The correct approach suggests itself when we think in Lagrangian terms following a freely-falling observer, whose worldline defines one of the threads. Such an observer feels no gravitational force at all but does notice that adjacent freely-falling observers do not necessarily move in straight lines with constant speed. In Newtonian terms this is explained by "tidal forces" while in general relativity it is called geodesic deviation. We shall not present a derivation of geodesic deviation here (one may find it in any general relativity textbook) but simply note that it follows from the non-commutativity of covariant spacetime derivatives of the 4-velocity. The relevant equation is the 4-dimensional version of the first of eqs. (4.5), called the Ricci identity:

$$[\boldsymbol{\nabla}_\kappa, \boldsymbol{\nabla}_\lambda]\, u^\mu = R^\mu{}_{\nu\kappa\lambda} u^\nu \ . \tag{4.74}$$

This identity holds for any differentiable vector field u^μ. *In the Lagrangian field approach we seek evolution equations for the Riemann tensor itself rather than the metric tensor components.*

One advantage of working with the Riemann tensor is the fact that part of it — the Ricci tensor — is given *algebraically* by the local stress-energy through eqs. (4.7) and (4.8). However, one cannot (in 4 dimensions) reconstruct the entire Riemann tensor from the Ricci tensor alone. One could obtain it by differentiating the metric found by solving the Einstein equations (cf. eqs. 4.9, 4.10). As we shall see, there is another method that does not require integrating the Einstein equations.

This alternative method is based on an evolution equation for that part of the Riemann tensor that cannot be obtained from the Ricci tensor, the Weyl tensor $C_{\mu\nu\kappa\lambda}$:

$$C_{\mu\nu\kappa\lambda} \equiv R_{\mu\nu\kappa\lambda} - \frac{1}{2}(g_{\mu\kappa}R_{\nu\lambda} + g_{\nu\lambda}R_{\mu\kappa} - g_{\mu\lambda}R_{\nu\kappa} - g_{\nu\kappa}R_{\mu\lambda}) + \frac{R}{6}(g_{\mu\kappa}g_{\nu\lambda} - g_{\mu\lambda}g_{\nu\kappa}) \ . \tag{4.75}$$

This tensor obeys all the symmetries of the Riemann tensor — $C_{\mu\nu\kappa\lambda} = C_{[\mu\nu][\kappa\lambda]} = C_{\kappa\lambda\mu\nu}$ and $C_{\mu[\nu\kappa\lambda]} = 0$ (where square brackets denote antisymmetrization) — and in addition is traceless: $C^{\kappa}{}_{\mu\kappa\nu} = 0$. Thus, the trace part of the Riemann tensor is given by the Ricci tensor $R_{\mu\nu}$ (through the Ricci terms on the right-hand side of eq. 4.75) while the traceless part is given by the Weyl tensor. Physically, the Ricci tensor gives the contribution to the spacetime curvature from local sources (through the Einstein eqs. 4.7 combined with 4.8) while the Weyl tensor gives the contribution due to nonlocal sources. It is clear that Newtonian tidal forces will be represented in the Weyl tensor. It may be shown that in 4 dimensions the Ricci and Weyl tensors each have 10 independent components.

How do we get an evolution equation for the Weyl tensor? The Einstein equations will not do because the Weyl tensor makes no appearance at all in the Einstein tensor. The correct method, due to Kundt & Trümper (1961), makes use of the Bianchi identities,

$$\boldsymbol{\nabla}_\sigma R_{\mu\nu\kappa\lambda} + \boldsymbol{\nabla}_\mu R_{\nu\sigma\kappa\lambda} + \boldsymbol{\nabla}_\nu R_{\sigma\mu\kappa\lambda} = 0 \ . \tag{4.76}$$

These identities follow directly from the definition of the Riemann tensor (see any general relativity or differential geometry textbook). For our purposes the key point is that they provide differential equations for the Riemann tensor. Contracting eq. (4.76) on κ and σ and using eqs. (4.75) and (4.8), we get

$$\boldsymbol{\nabla}^\kappa C_{\mu\nu\kappa\lambda} = \boldsymbol{\nabla}_{[\mu} G_{\nu]\lambda} + \frac{1}{3} g_{\lambda[\mu} \boldsymbol{\nabla}_{\nu]} G^\kappa{}_\kappa \ . \tag{4.77}$$

Note that if we contract now on λ and μ, using the symmetry of $G_{\mu\nu}$ and $g_{\mu\nu}$ we get $\boldsymbol{\nabla}_\mu G^\mu{}_\nu = 0$, as noted before. However, here we regard eq. (4.77) as an equation of motion for the Weyl tensor. Using the Einstein eqs. (4.7), we see that the source is given in terms of the energy-momentum tensor, so

$$\boldsymbol{\nabla}^\kappa C_{\mu\nu\kappa\lambda} = 8\pi G \left(\boldsymbol{\nabla}_{[\mu} T_{\nu]\lambda} + \frac{1}{3} g_{\lambda[\mu} \boldsymbol{\nabla}_{\nu]} T^\kappa{}_\kappa \right) \ . \tag{4.78}$$

The next step is to split the Weyl tensor into two second-rank tensors using a 1+3 threading of spacetime (Hawking 1966, Ellis 1971),

$$E_{\mu\nu}(u) \equiv u^\kappa u^\lambda C_{\mu\kappa\nu\lambda} \ , \quad H_{\mu\nu}(u) \equiv \frac{1}{2} \epsilon_{\alpha\beta\kappa(\mu} u^\kappa u^\lambda C^{\alpha\beta}{}_{\nu)\lambda} \ . \tag{4.79}$$

We have used the fully antisymmetric tensor $\epsilon_{\mu\nu\kappa\lambda} = (-g)^{1/2}\,[\mu\nu\kappa\lambda]$, where g is the determinant of $g_{\mu\nu}$ and $[\mu\nu\kappa\lambda]$ is the completely antisymmetric Levi-Civita symbol defined by three conditions: (1) $[0123] = +1$, (2) $[\mu\nu\kappa\lambda]$ changes sign if any two indices are exchanged, and (3) $[\mu\nu\kappa\lambda] = 0$ if any two indices are equal. (Note that Ellis uses the tensor $\eta_{\mu\nu\kappa\lambda} = -\epsilon_{\mu\nu\kappa\lambda}$. We have compensated for the sign change in defining $H_{\mu\nu}$. Beware that $\epsilon^{\mu\nu\kappa\lambda} = -(-g)^{-1/2}\,[\mu\nu\kappa\lambda]$.) The two new tensors $E_{\mu\nu}$ and $H_{\mu\nu}$ are both symmetric ($H_{\mu\nu}$ must be explicitly symmetrized), traceless, and flow-orthogonal, i.e., $E_{\mu\nu}u^\nu = H_{\mu\nu}u^\nu = 0$ and $P^\nu{}_\kappa E_{\mu\nu} = E_{\mu\kappa}$, $P^\nu{}_\kappa H_{\mu\nu} = H_{\mu\kappa}$. Therefore $E_{\mu\nu}$ and $H_{\mu\nu}$ each has 5 independent components, half as many as the Weyl tensor. Indeed, the Weyl tensor is fully determined by them for non-null threads:

$$\begin{aligned} C_{\mu\nu\kappa\lambda} = {} & (g_{\mu\nu\alpha\beta}\, g_{\kappa\lambda\gamma\delta} - \epsilon_{\mu\nu\alpha\beta}\, \epsilon_{\kappa\lambda\gamma\delta})\, u^\alpha u^\gamma E^{\beta\delta}(u) \\ & + (\epsilon_{\mu\nu\alpha\beta}\, g_{\kappa\lambda\gamma\delta} + g_{\mu\nu\alpha\beta}\, \epsilon_{\kappa\lambda\gamma\delta})\, u^\alpha u^\gamma H^{\beta\delta}(u) \ , \end{aligned} \tag{4.80}$$

where $g_{\mu\nu\alpha\beta} \equiv g_{\mu\alpha}g_{\nu\beta} - g_{\mu\beta}g_{\nu\alpha} = -\frac{1}{2}\epsilon_{\mu\nu}{}^{\kappa\lambda}\epsilon_{\kappa\lambda\alpha\beta} = g_{[\mu\nu][\alpha\beta]} = g_{\alpha\beta\mu\nu}$, with $g_{\mu[\nu\alpha\beta]} = 0$. Eq. (4.80) is the inverse of eqs. (4.79) provided $g_{\mu\nu}u^\mu u^\nu = \pm 1$. Ellis (1971) has a sign error in the first term of his version of eq. (4.80) at the end of his section 4.2.3.

The tensors $E_{\mu\nu}(u)$ and $H_{\mu\nu}(u)$ are called the electric and magnetic parts of the Weyl tensor, respectively. Together with the Ricci tensor they fully determine the spacetime curvature for a given threading (i.e., a system of threads with tangent vectors) $u^\mu(\lambda;\boldsymbol{q})$. It is worth noting that, if there are several threads at a given spacetime point, $E_{\mu\nu}(u)$ and $H_{\mu\nu}(u)$ have different values for each thread, and so they may be considered Lagrangian functions: $E_{\mu\nu}(\lambda;\boldsymbol{q})$ and $H_{\mu\nu}(\lambda;\boldsymbol{q})$. The Weyl tensor components are, however, unique, with the same value for all threads passing through the same spacetime point. This condition is satisfied automatically if the same 4-velocity u^μ is used in both eqs. (4.79) and (4.80).

Our goal is to rewrite eq. (4.78) in terms of $E_{\mu\nu}$ and $H_{\mu\nu}$. Because the results involve the covariant derivative of the 4-velocity field $\boldsymbol{\nabla}_\mu u_\nu$, we first decompose this quantity into acceleration, expansion, shear, and vorticity:

$$\begin{aligned} & \boldsymbol{\nabla}_\mu u_\nu = -u_\mu \frac{Du_\nu}{d\lambda} + P^\alpha{}_\mu P^\beta{}_\nu \boldsymbol{\nabla}_\alpha u_\beta = -u_\mu a_\nu + \frac{1}{3}\,\Theta P_{\mu\nu} + \sigma_{\mu\nu} + \omega_{\mu\nu}\,; \\ & \Theta = \boldsymbol{\nabla}_\mu u^\mu \ , \quad \sigma_{\mu\nu} = \sigma_{(\mu\nu)} \ , \quad \omega_{\mu\nu} = \omega_{[\mu\nu]} = \epsilon_{\mu\nu\alpha\beta}\, u^\alpha \omega^\beta \ . \end{aligned} \tag{4.81}$$

We have introduced the covariant derivative in the direction u^ν, $D/d\lambda \equiv u^\nu\boldsymbol{\nabla}_\nu$. Since this is just the proper time derivative along the worldline, $a_\nu = Du_\nu/d\lambda$ is the 4-acceleration. The flow-orthogonal part of the velocity gradient, $P^\alpha{}_\mu P^\beta{}_\nu \boldsymbol{\nabla}_\alpha u_\beta$, has been decomposed into the expansion scalar

Θ, the traceless shear tensor $\sigma_{\mu\nu}$, and the vorticity tensor $\omega_{\mu\nu}$ or its flow-orthogonal dual, ω^μ. Note that the expansion scalar includes a contribution due to cosmic expansion in addition to the peculiar velocity: neglecting metric perturbations, $\Theta = a^{-1}(\eta + \boldsymbol{\nabla}\cdot\boldsymbol{v})$. Note also that in the fluid rest frame, $\omega^i \boldsymbol{e}_i = \frac{1}{2}\boldsymbol{\nabla}\times\boldsymbol{v}$ is half the usual three-dimensional vorticity. (Ellis defines $\omega_{\mu\nu}$ and ω^μ with the opposite sign to us.)

We shall apply this gradient expansion to the tangent field of the 1+3 spacetime threading. This requires that u^μ be differentiable, which will be true (almost everywhere) if it corresponds to the 4-velocity field of a flow. In a frame comoving with the fluid, Θ, σ_{ij} and ω_{ij} are then the usual fluid expansion, shear, and vorticity, respectively.

By projecting $\boldsymbol{\nabla}^\kappa C_{\mu\nu\kappa\lambda}$ with various combinations of u^α and $P^{\alpha\beta}(u)$ (these are dependent on the spacetime threading), one can derive the following identities:

$$u^\nu u^\lambda \boldsymbol{\nabla}^\kappa C^\mu{}_{\nu\kappa\lambda} = P^\mu{}_\alpha P^\nu{}_\beta \boldsymbol{\nabla}_\nu E^{\alpha\beta} + \epsilon^{\mu\nu\alpha\beta}\, u_\nu \sigma_{\alpha\gamma} H^\gamma{}_\beta - 3H^\mu{}_\nu \omega^\nu \ , \tag{4.82}$$

$$\begin{aligned}\frac{1}{2}P^\mu{}_\alpha u_\beta u^\lambda \epsilon^{\alpha\beta\gamma\delta}\boldsymbol{\nabla}^\kappa C_{\gamma\delta\kappa\lambda} = -P^\mu{}_\alpha P^\nu{}_\beta \boldsymbol{\nabla}_\nu H^{\alpha\beta} &+ \epsilon^{\mu\nu\alpha\beta}\, u_\nu \sigma_{\alpha\gamma} E^\gamma{}_\beta \\ &-3E^\mu{}_\nu \omega^\nu \ ,\end{aligned} \tag{4.83}$$

$$\begin{aligned}P^{\mu\lambda}P^{\nu\alpha}u^\beta \boldsymbol{\nabla}^\kappa C_{\alpha\beta\kappa\lambda} \ = \ & P^\mu{}_\alpha P^\nu{}_\beta \frac{DE^{\alpha\beta}}{d\lambda} + P^{\alpha\nu}\epsilon^{\mu\beta\gamma\delta}u_\beta \boldsymbol{\nabla}_\gamma H_{\alpha\delta} \\ & +2u_\alpha a_\beta H_\gamma{}^{(\mu}\epsilon^{\nu)\alpha\beta\gamma} + \Theta E^{\mu\nu} + P^{\mu\nu}(\sigma^{\alpha\beta}E_{\alpha\beta}) \\ & -2E^{\alpha\nu}(\sigma^\mu{}_\alpha - \omega^\mu{}_\alpha) - E^{\alpha\mu}(\sigma^\nu{}_\alpha + \omega^\nu{}_\alpha) \ ,\end{aligned} \tag{4.84}$$

$$\begin{aligned}\frac{1}{2}P^\mu{}_\alpha P^{\nu\lambda}u_\beta \epsilon^{\alpha\beta\gamma\delta}\boldsymbol{\nabla}^\kappa C_{\gamma\delta\kappa\lambda} = & -P^\mu{}_\alpha P^\nu{}_\beta \frac{DH^{\alpha\beta}}{d\lambda} + P^{\alpha\mu}\epsilon^{\nu\beta\gamma\delta}u_\beta \boldsymbol{\nabla}_\gamma E_{\alpha\delta} \\ & +2u_\alpha a_\beta E_\gamma{}^{(\mu}\epsilon^{\nu)\alpha\beta\gamma} - \Theta H^{\mu\nu} - P^{\mu\nu}(\sigma^{\alpha\beta}H_{\alpha\beta}) \\ & +2H^{\alpha\mu}(\sigma^\nu{}_\alpha - \omega^\nu{}_\alpha) + H^{\alpha\nu}(\sigma^\mu{}_\alpha + \omega^\mu{}_\alpha) \ .\end{aligned} \tag{4.85}$$

These identities follow from eqs. (4.80) and (4.81). All quantities on the right-hand sides are to be evaluated for a given thread $u^\mu(\lambda;\boldsymbol{q})$.

Finally we are ready to obtain equations of motion for the electric and magnetic parts of the Weyl tensor from eq. (4.78). In fact, infinitely many sets of equations are possible because are free to use any spacetime threading! For example, we may choose *Eulerian* threading with $\boldsymbol{q} = \boldsymbol{x}$, in which case in the Poisson gauge we have $u^0 = a^{-1}(1-\psi)$ and $u^i = 0$, so that $D/d\lambda = a^{-1}(1-\psi)\partial_\tau$ is the Eulerian proper time derivative. In this case the 1+3 split coincides with our previous 3+1 split. The Eulerian description is not covariant, for it depends on our choice of gauge. Because the Weyl tensor formalism is more complicated than our previous treatment based on the Einstein equations, there is no clear advantage to its use with Eulerian threading.

If, however, we use the fluid velocity itself — the u^μ appearing in eq. (4.19), which is well-defined even for an imperfect or collisionless fluid — to define the threading, then the Weyl tensor approach becomes more attractive. This choice corresponds to *Lagrangian* threading: the threads are the worldlines of fluid elements, so that $D/d\lambda$ now is the proper time derivative measured in the fluid rest frame. There are two important advantages to this choice. First, it is covariant: the fluid worldlines define a unique spacetime threading with no gauge ambiguities (Ellis & Bruni 1989), while any coordinates may be used to express the tensor components $E_{\mu\nu}$ and $H_{\mu\nu}$. Second, the right-hand side of eq. (4.78) — the source for the Weyl tensor — is expressed in terms of the same 4-velocity used in the threading, greatly simplifying the projections appearing in eqs. (4.82)–(4.85).

Ellis (1971) and Hwang & Vishniac (1990) give the Lagrangian gravitational field equations for a general stress-energy tensor. For a perfect fluid (with $\Sigma^{\mu\nu}=0$ in eq. 4.19) the results are

$$\begin{aligned}
(\text{div-}E):\quad & P^\mu{}_\alpha P^\nu{}_\beta \boldsymbol{\nabla}_\nu E^{\alpha\beta} + \epsilon^{\mu\nu\alpha\beta}u_\nu \sigma_{\alpha\gamma}H^\gamma{}_\beta - 3H^\mu{}_\nu\omega^\nu \\
& \qquad = \frac{8\pi}{3}GP^{\mu\nu}\boldsymbol{\nabla}_\nu\rho \ , \qquad (4.86)\\
(\dot{H}):\quad & P^\mu{}_\alpha P^\nu{}_\beta \frac{DH^{\alpha\beta}}{d\lambda} - P^{\alpha(\mu}\epsilon^{\nu)\beta\gamma\delta}u_\beta\boldsymbol{\nabla}_\gamma E_{\alpha\delta} \\
& -2u_\alpha a_\beta E_\gamma{}^{(\mu}\epsilon^{\nu)\alpha\beta\gamma} + \Theta H^{\mu\nu} + P^{\mu\nu}(\sigma^{\alpha\beta}H_{\alpha\beta}) - 3H^{\alpha(\mu}\sigma^{\nu)}{}_\alpha \\
& \qquad + H^{\alpha(\mu}\omega^{\nu)}{}_\alpha = 0 \ , \qquad (4.87)\\
(\text{div-}H):\quad & P^\mu{}_\alpha P^\nu{}_\beta \boldsymbol{\nabla}_\nu H^{\alpha\beta} - \epsilon^{\mu\nu\alpha\beta}u_\nu \sigma_{\alpha\gamma}E^\gamma{}_\beta + 3E^\mu{}_\nu\omega^\nu \\
& \qquad = -8\pi G(\rho+p)\omega^\mu \ , \qquad (4.88)\\
(\dot{E}):\quad & P^\mu{}_\alpha P^\nu{}_\beta \frac{DE^{\alpha\beta}}{d\lambda} + P^{\alpha(\mu}\epsilon^{\nu)\beta\gamma\delta}u_\beta\boldsymbol{\nabla}_\gamma H_{\alpha\delta} \\
& +2u_\alpha a_\beta H_\gamma{}^{(\mu}\epsilon^{\nu)\alpha\beta\gamma} + \Theta E^{\mu\nu} + P^{\mu\nu}(\sigma^{\alpha\beta}E_{\alpha\beta}) - 3E^{\alpha(\mu}\sigma^{\nu)}{}_\alpha \\
& \qquad + E^{\alpha(\mu}\omega^{\nu)}{}_\alpha = -4\pi G(\rho+p)\sigma^{\mu\nu} \ . \qquad (4.89)
\end{aligned}$$

These have been obtained by substituting eqs. (4.19) and (4.82)–(4.85) into eq. (4.78), and using $\boldsymbol{\nabla}_\nu T^{\mu\nu}=0$ to simplify the right-hand sides of the div-E and $\dot{E}$ equations. The results agree with eqs. (4.21) of Ellis

(1971). For an imperfect fluid it is necessary to add terms to the right-hand sides involving the shear stress $\Sigma^{\mu\nu}$. For a pressureless fluid (e.g., cold dust before the intersection of trajectories) the 4-acceleration a_β vanishes.

In his beautifully lucid pedagogical articles presenting the Lagrangian fluid approach, Ellis (1971, 1973) has noted the similarity of eqs. (4.86)–(4.89) to the Maxwell equations, particularly if the covariant form of the latter are split using 1+3 threading. Compare them with eqs. (4.62) for the vector (not tensor) gravitational fields in the Poisson gauge. Although the latter equations are more reminiscent of the Maxwell equations in flat spacetime, they are only approximate (they are based on a linearized metric and neglect several generally small terms), they are tied to a particular coordinate system (Poisson gauge), and they do not incorporate gravitational radiation. By contrast, eqs. (4.86)–(4.89) are exact, they are valid in any coordinate system (all quantities appearing in them are spacetime tensors), and they include all gravitational effects. The exact equations involve second-rank tensors rather than vectors because, in the terminology of particle physics, gravity is a spin-2 rather than a spin-1 gauge theory.

The quasi-Maxwellian equations (4.86)–(4.89) show that the evolution of the Weyl tensor depends on the fluid velocity gradient. This quantity could be computed by evolving the equations of motion for the matter (e.g., eqs. 4.24 and 4.25) to get the velocity field $u^\mu(x)$ and then taking its derivatives. However, there is a more natural way in the context of the Lagrangian approach: integrate evolution equations for the velocity gradient itself. In fact, such equations follow simply from projecting the Ricci identity (4.74) for the fluid velocity u^μ with $u^\kappa P^{\alpha\lambda}P_{\beta\mu}$ and separating the result as in eqs. (4.81). It is straightforward to derive the following equations (Ellis 1971, 1973):

$$\frac{D\Theta}{d\lambda} - \boldsymbol{\nabla}_\mu a^\mu + \frac{1}{3}\,\Theta^2 + \sigma^{\mu\nu}\sigma_{\mu\nu} - 2\omega^2 = -4\pi G(\rho + 3p) \ , \tag{4.90}$$

$$P^\mu{}_\nu \frac{D\omega^\nu}{d\lambda} + \frac{1}{2}\,\epsilon^{\mu\nu\alpha\beta}u_\nu \boldsymbol{\nabla}_\alpha a_\beta + \frac{2}{3}\,\Theta\omega^\mu - \sigma^\mu{}_\nu\omega^\nu = 0 \ , \tag{4.91}$$

$$\begin{aligned} &P^\mu{}_\alpha P^\nu{}_\beta \frac{D\sigma^{\alpha\beta}}{d\lambda} - \boldsymbol{\nabla}^{(\mu}a^{\nu)} + \frac{2}{3}\,\Theta\sigma^{\mu\nu} + \sigma^{\mu\alpha}\sigma^\nu{}_\alpha + \omega^\mu\omega^\nu \\ &-\frac{1}{3}\,P^{\mu\nu}\left(\sigma^{\alpha\beta}\sigma_{\alpha\beta} + \omega^2 - \boldsymbol{\nabla}_\alpha a^\alpha\right) = -E^{\mu\nu} \ , \end{aligned} \tag{4.92}$$

where $\omega^2 \equiv \omega^\mu\omega_\mu$. Equation (4.90) is known as the Raychaudhuri equation. It shows that the expansion is decelerated by the shear and by the local density and pressure (if $\rho + 3p > 0$), but is accelerated by the vorticity. Vorticity, on the other hand, is unaffected by gravity; eq. (4.91) implies that vorticity can be described by field lines that (if a^μ vanishes or if the fluid has vanishing shear stress) are frozen into the fluid (Ellis 1973). Finally, shear, being the traceless symmetric part of the velocity gradient tensor, has as its source the electric part of the Weyl tensor. These equations are essentially identical to their Newtonian counterparts (Ellis 1971; Bertschinger & Jain 1994). Note that the magnetic part of the Weyl tensor does not directly influence the matter evolution.

Closing the Lagrangian field equations also requires specifying the evolution of density and pressure (and shear stress, if present). These follow from energy conservation, $\boldsymbol{\nabla}_\nu T^{\mu\nu} = 0$, combined with an equation of state. For a perfect fluid, using eq. (4.19) with $\Sigma^{\mu\nu} = 0$ and projecting the divergence of the stress-energy tensor with u_μ gives

$$\frac{D\rho}{d\lambda} + (\rho + p)\Theta = 0 \ . \tag{4.93}$$

Equations (4.86)–(4.93) now provide a set of Lagrangian equations of motion for the matter and spacetime curvature variables following a mass element. These Lagrangian equations of motion offer a powerful approach to general relativity — and to relativistic cosmology and perturbation theory — that is quite different from the usual methods based on integration of the Einstein equations in a particular gauge (or with gauge-invariant variables).

To relate the relativistic Lagrangian approach to dynamics to the standard Newtonian one, we now evaluate the electric and magnetic parts of the Weyl tensor in the weak-field, slow-motion limit. They involve second derivatives of the metric and not simply the first derivatives present in eqs. (4.61). In the Poisson gauge, to lowest order in the metric perturbations and the velocity, from eqs. (4.79) one obtains (Bertschinger & Hamilton 1994)

$$\begin{aligned} E_{ij} &= \frac{1}{2}\,D_{ij}(\psi + \phi) + \frac{1}{2}\,\boldsymbol{\nabla}_{(i}\dot{w}_{j)} - \frac{1}{2}\left(\ddot{h}_{ij} + \boldsymbol{\nabla}^2 h_{ij} - 2K h_{ij}\right) \ , \\ H_{ij} &= -\frac{1}{2}\,\boldsymbol{\nabla}_{(i}H_{j)} + \epsilon_{kl(i}\boldsymbol{\nabla}^k \dot{h}_{j)}{}^l \ , \end{aligned} \tag{4.94}$$

where H_j is the gravitomagnetic field defined in eq. (4.61). The time-time and space-time components of $E_{\mu\nu}$ and $H_{\mu\nu}$ vanish in the fluid frame because these tensors are flow-orthogonal.

Do these results imply that in the Newtonian limit $H_{ij} = 0$ and $E_{ij} = D_{ij}\phi$ is simply the gravitational tidal field? If we say that the Newtonian limit implies $\psi = \phi$ and $w_i = h_{ij} = 0$ (no relativistic shear stress, no gravitomagnetism, and no gravitational radiation), then the answer would appear to be yes. This possibility, considered by Matarrese, Pantano, & Saez (1993) and Bertschinger & Jain (1994), has an important implication: for cold dust, the Lagrangian evolution of the tidal tensor obtained from eq. (4.89) would then be purely local (Barnes & Rowlingson 1989). That is, the evolution of the tide (the electric part of the Weyl tensor) along the thread $u^\mu(\lambda;\boldsymbol{q})$ would depend only on the density, velocity gradient, and tide defined at each point along the trajectory with no further spatial gradients (since they arise only from the magnetic terms in eq. 4.89). The evolution of the density and of the velocity gradient tensor are clearly local (eqs. 4.90–4.93, with $a^\mu = 0$) aside from the tidal tensor, but we have just seen that its evolution depends only on other local quantities. In other words, if $H_{ij} = 0$, the matter and spacetime curvature variables would evolve independently along different fluid worldlines. Bruni, Matarrese, and Pantano (1994) call this a "silent universe."

Local evolution does occur if the metric perturbations are one-dimensional (e.g., the Bondi-Tolman solution in spherical symmetry, or the Zel'dovich solution in plane symmetry; see Matarrese et al. 1993 and Croudace et al. 1994), but it would be surprising were this to happen for arbitrary matter distributions in the Newtonian limit.

Bertschinger & Hamilton (1994) and Kofman & Pogosyan (1995) have shown that, in fact, the general evolution of the tidal tensor in the Newtonian limit is nonlocal. The reason is that, while one may neglect the metric perturbation w_i in the Newtonian limit, its gradient should not be neglected. Doing so violates the transverse momentum constraint equation (4.51), unless the transverse momentum density (the source term for $\boldsymbol{w}$ in the Poisson gauge) vanishes. This condition does not hold for general motion in the Newtonian limit.

A convincing proof of nonlocality is given by the derivation of eq. (4.89) in locally flat coordinates in the fluid frame by Bertschinger & Hamilton (1994) using only the Newtonian continuity and Poisson equations plus the second pair of eqs. (4.62) and a modified form of eq. (4.94):

$$H_{ij} = -\frac{1}{2}\,\boldsymbol{\nabla}_{(i}H_{j)} - 2v_k\epsilon^{kl}{}_{(i}E_{j)l} + O(v/c)^2 \ . \tag{4.95}$$

This is taken as the definition of H_{ij} in the Newtonian limit (where we also have $E_{ij} = D_{ij}\phi$). Note that in the Newtonian limit we neglect gravitational radiation, but we must include terms that are first-order in the velocity. Even though we define the magnetic part of the Weyl tensor using the fluid 4-velocity, we are evaluating its components in a particular gauge — Poisson gauge — in which the 3-velocity does not necessarily vanish. The extra term in eq. (4.95) arises from evaluating eqs. (4.79) to first order in v/c (Bertschinger & Hamilton 1994) and it is analogous to the Lorentz transformation of electric fields into magnetic fields in a moving frame. Both terms in eq. (4.95) are of order $G\rho\boldsymbol{v}$. They can not be neglected in the Newtonian limit.

The implication of this result is that Lagrangian evolution of matter and gravity is not purely local except under severe restrictions such as spherical or plane symmetry. There exist, of course, local approximations to evolution such as the Zel'dovich (1970) approximation. Finding improved local approximations is one of the active areas of research in large-scale structure theory. Formulating the problem in terms of the Lagrangian fluid and field equations not only may suggest new approaches, it is also likely to clarify the relation between general relativity and Newtonian dynamics.

References

[1] L. Abbott & D. Harari, Nucl. Phys. B264 (1986) 487.
[2] R. Arnowitt, S. Deser & C. W. Misner, Gravitation, in: ed. L. Witten (Wiley, New York, 1972) p. 227.
[3] J. M. Bardeen, Phys. Rev. D22 (1980) 1882.
[4] J. M. Bardeen, Particle Physics and Cosmology, in: ed. A. Zee (Gordon and Breach, New York, 1989).
[5] A. Barnes & R. R. Rowlingson, Class. Quant. Grav. 6 (1989) 949.
[6] E. Bertschinger, Statistical Description of Transport in Plasma, Astro-, and Nuclear Physics, in: eds. J. Misguich, G. Pelletier & P. Schuck (Nova Science, Commack, NY, 1993), p. 193.
[7] E. Bertschinger & A. Dekel, Ap. J. Lett. 336 (1989) L5.
[8] E. Bertschinger & J. M. Gelb, Comput. Phys. 5 (1991) 164.
[9] E. Bertschinger & A. J. S. Hamilton, Ap. J. 435 (1994) 1.
[10] E. Bertschinger & B. Jain, Ap. J. 431 (1994) 486.
[11] E. Bertschinger & P. N. Watts, Ap. J. 328 (1988) 23.
[12] J. Binney & S. Tremaine, Galactic Dynamics (Princeton Univ. Press, Princeton, 1987).
[13] L. Bombelli, W. E. Couch & R. J. Torrence, Class. Quant. Grav. 11 (1994) 139.
[14] J. R. Bond & G. Efstathiou, M.N.R.A.S. 226 (1987) 665.
[15] J. R. Bond & A. S. Szalay, Ap. J. 274 (1983) 443.
[16] O. L. Brill & B. Goodman, Am. J. Phys. 35 (1967) 832.
[17] M. Bruni, P. K. S. Dunsby & G. F. R. Ellis, Ap. J. 395 (1992) 34.
[18] M. Bruni, S. Matarrese & O. Pantano (1994) preprint astro-ph/9406068.
[19] G. L. Bryan, R. Cen, M. L. Norman, J.P. Ostriker & J. M. Stone, Ap. J. 428 (1994) 405.
[20] K. M. Croudace, J. Parry, D. S. Salopek & J. M. Stewart, Ap. J. 423 (1994) 22.
[21] A. Dekel, Ann. Rev. Astron. Ap. 32 (1994) 371.

[22] A. Dekel, E. Bertschinger, A. Yahil, M. Strauss, M. Davis & J. Huchra, Ap. J. 412 (1993) 1.
[23] R. Durrer & N. Straumann, Helv. Phys. Acta 61 (1988) 1027.
[24] G. Efstathiou, Physics of the Early Universe, in: eds. J. A. Peacock, A. F. Heavens & A. T. Davies (IOP, Bristol, 1990) p. 361.
[25] G. Efstathiou & J. R. Bond, M.N.R.A.S. 218 (1986) 103.
[26] G. Efstathiou, M. Davis, C. S. Frenk & S. D. M. White, Ap. J. Suppl. 57 (1985) 241.
[27] J. Ehlers, Akad. Wiss. Lit. Mainz Abh. Math.-Nat. Kl. 11 (1961).
[28] J. Ehlers, General Relativity and Cosmology, in: ed. R. K. Sachs (Academic Press, New York, 1971) p. 1.
[29] A. Einstein, Preuss. Akad. Wiss. Berlin, Sitzber. (1917) 142.
[30] G. F. R. Ellis, General Relativity and Cosmology, in: ed. R. K. Sachs (Academic Press, New York, 1971), p. 104.
[31] G. F. R. Ellis, Cargèse Lectures in Physics, vol. 6, in: ed. E. Schatzman (Gordon and Breach, New York, 1973), p. 1.
[32] G. F. R. Ellis & M. Bruni, Phys. Rev. D40 (1989) 1804.
[33] I. H. Gilbert, Ap. J. 144 (1966) 233.
[34] H. Goldstein, Classical Mechanics (Addison-Wesley, Reading, 1980).
[35] S. Hawking Ap. J. 145 (1966) 544.
[36] R. W. Hockney & J. W. Eastwood, Computer Simulation Using Particles (McGraw-Hill, New York, 1981).
[37] J. A. Holtzman, Ap. J. Suppl. 71 (1989) 1.
[38] J.-C. Hwang & E. T. Vishniac, Ap. J. 353 (1990) 1.
[39] S. Ichimaru, Statistical Plasma Physics (Addison-Wesley, Redwood City, 1992).
[40] W. M. Irvine, Ann. Phys. (New York) 32 (1965) 322.
[41] R. T. Jantzen, P. Carini & D. Bini, Ann. Phys. 215 (1992) 1.
[42] J. H. Jeans, Phil. Trans. 199A (1902) 1.
[43] H. Kang, J. P. Ostriker, R. Cen, D. Ryu, L. Hernquist, A. E. Evrard, G. L. Bryan & M. L. Norman, Ap. J. 430 (1994) 83.
[44] Yu. L. Klimontovich, The Statistical Theory of Non-Equilibrium Processes in a Plasma (MIT Press, Cambridge, 1967).
[45] H. Kodama & M. Sasaki, Prog. Theor. Phys. Suppl. 78 (1984) 1.
[46] H. Kodama & M. Sasaki, Int. J. Mod. Phys. A1 (1986) 265; A2 (1986) 491.
[47] L. Kofman & D. Pogosyan, Ap. J. 442 (1995) 30.
[48] E. W. Kolb & M. S. Turner, The Early Universe (Addison-Wesley, Redwood City, 1990).
[49] W. Kundt & M. Trümper, Akad. Wiss. Lit. Mainz Abh. Math.-Nat. Kl. 12 (1961).
[50] O. Lahav, P. B. Lilje, J. R. Primack & M. J. Rees, M.N.R.A.S. 251 (1991) 128.
[51] L. D. Landau & E. M. Lifshitz, Fluid Mechanics (Pergamon Press, Oxford, 1959).
[52] D. Layzer, Ap. J. 138 (1963) 174.
[53] R. J. Leveque, Numerical methods for conservation laws (Birkhauser, Boston, 1992).
[54] A. R. Liddle & D. H. Lyth, Phys. Rep. 231 (1993) 1.
[55] E. M. Lifshitz, J. Phys. (USSR) 10 (1946) 116.
[56] E. M. Lifshitz & I. M. Khalatnikov, Adv. Phys. 12 (1963) 185.
[57] D. H. Lyth & E. D. Stewart, Ap. J. 361 (1990) 343.
[58] C.-P. Ma & E. Bertschinger, Ap. J. 429 (1994a) 22.
[59] C.-P. Ma & E. Bertschinger, (1994b) preprint astro-ph/9401007.
[60] S. Matarrese, O. Pantano & D. Saez, Phys. Rev. D47 (1993) 1311.

[61] C. W. Misner, K. S. Thorne & J. A. Wheeler, Gravitation (Freeman, San Francisco, 1973).
[62] J. J. Monaghan, Ann. Rev. Astron. Ap. 30 (1992) 543.
[63] V. F. Mukhanov, H. A. Feldman & R. H. Brandenberger, Phys. Rep. 215 (1992) 1.
[64] M. K. Munitz, Theories of the Universe (Free Press, New York, 1957).
[65] T. Padmanabhan, Structure Formation in the Universe (Cambridge Univ. Press, Cambridge, 1993).
[66] J. Pedolsky, Geophysical Fluid Dynamics (Springer-Verlag, New York, 1987).
[67] P. J. E. Peebles, The Large-Scale Structure of the Universe (Princeton Univ. Press, Princeton, 1980).
[68] P. J. E. Peebles, Principles of Physical Cosmology (Princeton Univ. Press, Princeton, 1993).
[69] P. J. E. Peebles & J. T. Yu, Ap. J. 162 (1970) 815.
[70] W. H. Press & E. T. Vishniac, Ap. J. 239 (1980) 1.
[71] C. Pryor & J. Kormendy, Astron. J. 100 (1990) 127.
[72] B. Ratra, Phys. Rev. D38 (1988) 2399.
[73] A. Rebhan, Nucl. Phys. B369 (1992) 479.
[74] D. S. Salopek & J. M. Stewart, Class. Quant. Grav. 9 (1992) 1943.
[75] B. Schutz, A First Course in General Relativity (Cambridge Univ. Press, Cambridge, 1985).
[76] S. Setayeshgar, SB thesis (MIT, 1990).
[77] G. A. Sod, Numerical Methods in Fluid Dynamics (Cambridge Univ. Press, Cambridge, 1985).
[78] J. M. Stewart, Non-Equilibrium Relativistic Kinetic Theory, lecture notes in Physics 10 (Springer-Verlag, Berlin, 1971).
[79] J. M. Stewart, Class. Quant. Grav. 7 (1990) 1169.
[80] K. S. Thorne, R. H. Price & D. A. Macdonald, Black Holes: The Membrane Paradigm (Yale Univ. Press, New Haven, 1986).
[81] S. Tremaine & J. E. Gunn, Phys. Rev. Lett. 42 (1979) 407.
[82] S. Weinberg, Gravitation and Cosmology (Wiley, New York, 1972).
[83] M. L. Wilson, Ap. J. 273 (1983) 2.
[84] M. L. Wilson & J. Silk, Ap. J. 243 (1981) 14.
[85] Ya. B. Zel'dovich, Astron. Ap. 5 (1970) 84.
[86] Ya. B. Zel'dovich & I. D. Novikov, The Structure and Evolution of the Universe (Univ. of Chicago Press, Chicago, 1983).