Mon. Not. R. Astron. Soc. **264,** 729–739 (1993)

Relativistic hydrodynamics and gravitational instability revisited

J. C. Jackson
Department of Mathematics and Statistics, University of Northumbria at Newcastle, Ellison Building, Newcastle upon Tyne, NE1 8ST

Accepted 1993 April 19. Received 1993 March 4

ABSTRACT
After some 90 years of effort, the phenomenon of gravitational instability elucidated by Jeans continues to attract much attention and to reveal an unexpected richness of structure. A unified and simplified treatment of recent development in this area is presented, and extended to give new, completely general, exact second-order acoustic propagation equations, governing the evolution of the comoving fractional spatial gradient of the density. Newtonian theory and general relativity are considered, and the unity of the two points of view is emphasized. The Einstein static universe and expanding homogeneous isotropic cosmological models are re-examined.

Key words: gravitation – hydrodynamics – instabilities – relativity – large-scale structure of Universe.

1 INTRODUCTION

An extensive medium tends to fragment under the action of gravity, which phenomenon was first put on a quantitative footing by Jeans (1902) in his celebrated theorem. This footing was not entirely sound, as the medium was assumed to be static and infinite in extent, and to comprise a homogeneous, compressible perfect fluid. Jeans considered the propagation of an acoustic disturbance, governed by the basic equations of Newtonian hydrodynamics and gravitation, linearized with respect to the aforesaid static background. We now know that a non-static homogeneous isotropic background would have been more appropriate, corresponding to the Newtonian version of a Friedmann–Lemaître–Robertson–Walker (FLRW) universe. Such a consistent Newtonian picture did not emerge (McCrea & Milne 1934; McCrea 1955; Heckmann & Schücking 1955, 1956; Bonner 1957) until long after the advent of Newtonian cosmology, and benefited to some extent from hindsight, arising from the early development of this topic within the framework of relativistic cosmology (Lifshitz 1946; Lifshitz & Khalatnikov 1963).

The setting for Newtonian calculations is of course fixed Galilean space and time, whereas in general relativity space–time itself is part of the dynamics. For this reason, early relativistic work concentrated upon perturbations of the metric and the corresponding Einstein equations, and looked very different from the corresponding Newtonian view, for which the starting points are the principles of conservation of mass and momentum in the form of the equation of continuity and Euler's equation of motion. The two points of view were thus difficult to compare, and the simpler Newtonian approach provided little guidance for further relativistic endeavours. An influential advance in this context came with the work of Hawking (1966), who used the corresponding relativistic conservation laws, which are a direct consequence of Einstein's equations (via the Ricci identities). These were linearized about a FLRW solution, with a fixed Robertson–Walker metric, that is, metric perturbations were (erroneously) ignored. Following the work of Hawking, I conceived the notion of an exact second-order equation (Jackson 1972), which I called the general relativistic equation of finite-amplitude sound propagation, to be derived from the appropriate first-order equations without any approximation. This turned out to be the relativistic analogue of a Newtonian equation derived some years earlier by Hunter (1964). The advantage of such a master equation is that it governs the evolution of any space–time, including one with high symmetry, in which covariantly defined elements of the fluid flow are exactly zero; in a space–time that is close to the latter, the said elements will be small, and appropriate non-linear terms in the master equation can be neglected. As a simple illustration of this philosophy I considered the Einstein static universe, as an updated version of Jeans's original concept, to obtain a relativistic version of his instability criterion. [This has been described as being in error by Ellis, Hwang & Bruni (1989) and Hwang & Vishniac (1990), and incompatible with their results; a very minor purpose of this work is to effect a reconciliation in this context.]

Despite its long and distinguished history, the subject continues to attract much attention, and to reveal an unexpected richness of structure. Owing to the fully covariant nature of general relativity, a given physical situation can appear in various guises, related by coordinate transformations. The consequent problems were first addressed fully by Bardeen (1980), who observed that 'if the gauge condition imposed to simplify the metric leaves a residual gauge freedom, the perturbation equation will have spurious gauge mode solutions which can be completely annulled by a gauge transformation and have no physical reality'. Disentangling of these modes from those with real physical content has proved to be a major undertaking. Bardeen adopted the 'old fashioned' approach, concentrating on the perturbed Einstein equations, and identified a set of perturbed quantities which are invariant with respect to the residual infinitesimal coordinate freedom; in other words, they take the same value in all nearby coordinate systems (Stewart 1990). Although Bardeen's work brought gauge problems into sharp focus, subsequent endeavours have concentrated on simpler realization of his ideas. Highlights are perhaps Hwang & Vishniac's (1990) completion of the Hawking–Jackson programme, with due allowance for changes in geometry, and particularly a series of papers by Ellis and his co-workers (Ellis & Bruni 1989; Ellis et al. 1989; Ellis, Bruni & Hwang 1990; Ellis 1990), who introduced the comoving fractional spatial gradient of the energy density as the proper gauge-invariant quantity in this context.

The purpose of this paper is to present a simplified and unified account of recent developments in this area, with particular reference to the work of Ellis et al., and to my own earlier work (Jackson 1972). I shall concentrate on a perfect fluid with density μ and pressure p related by a barotropic equation of state

$$p=p(\mu), \tag{1}$$

and $c_s=(\mathrm{d}p/\mathrm{d}\mu)^{1/2}$ will occasionally denote the velocity of sound in the medium, when $\mathrm{d}p/\mathrm{d}\mu>0$. I shall be concerned with exact equations, particularly exact second-order propagation equations, in which respect the work goes further than its predecessors. My main purpose is to establish these equations, rather than to consider new applications. Nevertheless, the homogeneous isotropic case will be re-examined, and some results which are well-known (or at least implicit in earlier work) will be re-stated. New applications might include an examination of non-linear effects in homogeneous isotropic models, and linear instability in models with less symmetry, i.e. those that are homogeneous but anisotropic.

2 NOTATION AND CONVENTIONS

Greek indices will refer to Galilean space coordinates, and ∇_α will indicate the standard spatial gradient with respect to these coordinates. Thus $\nabla_\alpha\mu=\mu_{;\alpha}$ and $\nabla_\alpha u_\beta=u_{\beta;\alpha}$, where a semi-colon means covariant differentiation with respect to these coordinates, and u_α is the fluid flow vector. A dot will indicate the total (Lagrangian) time derivative, thus $\dot{\mu}=\partial\mu/\partial t+\mu_{\alpha;\beta}u^\beta$. The gradient of the fluid flow vector is decomposed as follows:

$$u_{\alpha;\beta}=\omega_{\alpha\beta}+\sigma_{\alpha\beta}+\tfrac{1}{3}\theta g_{\alpha\beta} \tag{2}$$

where the vorticity tensor $\omega_{\alpha\beta}=u_{[\alpha;\beta]}$, the shear tensor $\sigma_{\alpha\beta}=u_{(\alpha;\beta)}-\frac{1}{3}\theta g_{\alpha\beta}$, and the expansion $\theta=u^\alpha{}_{;\alpha}$; square and round brackets denote anti-symmetrization and symmetrization respectively, and $g_{\alpha\beta}$ is the Galilean space metric. (Although Newtonian theory will be couched in the language of tensor calculus, Galilean space is of course flat; covariant derivatives commute: $u_{\alpha;[\beta\gamma]}=0$, and there would be no loss of generality in using Cartesian coordinates and ordinary derivatives.)

Italic indices will refer to space–time coordinates, and g_{ab} is now the space–time metric. ∇_a will indicate the *spatial* gradient with respect to thse coordinates, that is

$$\nabla_a\mu=h_a{}^b\mu_{;b} \tag{3}$$

where $h_{ab}=g_{ab}+u_au_b$ projects into the three-space orthogonal to the time-like vector u_a, in this case the normalized fluid four-velocity ($u_au^a=-1$). [The reader should note that my use of ∇_a here is non-standard, as it usually means the ordinary covariant derivative; the operator defined by equation (3) is often denoted by ${}^{(3)}\nabla_\alpha$; the equations to be presented here are much easier to read if the ${}^{(3)}$ is omitted.] The first covariant derivative of u^a is now

$$u_{a;b}=\omega_{ab}+\sigma_{ab}+\tfrac{1}{3}\theta h_{ab}-\dot{u}_au_b \tag{4}$$

where $\omega_{ab}=h_a{}^ch_b{}^du_{[c;d]}$, $\sigma_{ab}=h_a{}^ch_b{}^du_{(c;d)}-\frac{1}{3}\theta h_{ab}$, $\theta=u^a{}_{;a}$, and $\dot{u}_a=u_{a;b}u^b$; the corresponding spatial gradient is

$$\nabla_bu_a=h_a{}^ch_b{}^du_{c;d}=\omega_{ab}+\sigma_{ab}+\tfrac{1}{3}\theta h_{ab}. \tag{5}$$

The Riemann tensor is defined by

$$u_{a;bc}-u_{a;cb}=R^d{}_{abc}u_d \tag{6}$$

and the Ricci tensor by $R_{ab}=R^c{}_{acb}$. With these conventions, Einstein's equations are

$$R_{ab}-\tfrac{1}{2}Rg_{ab}+\Lambda g_{ab}=\kappa T_{ab} \tag{7}$$

where Λ is the cosmological constant, and $\kappa=8\pi G$, in units in which $c=1$; T_{ab} is the energy–momentum tensor, which for a perfect fluid has the form

$$T_{ab}=(\mu+p)u_au_b+pg_{ab}. \tag{8}$$

3 NEWTONIAN EQUATIONS

We start with the equation of continuity

$$\frac{\dot{\mu}}{\mu}+\theta=0 \tag{9}$$

and Euler's equation of motion

$$\dot{u}_\alpha+\frac{1}{\mu}\nabla_\alpha p+\nabla_\alpha\Phi=0 \tag{10}$$

where Φ is the gravitational potential, governed by Poisson's equation

$$\nabla^2\Phi=4\pi G\mu-\Lambda. \tag{11}$$

If we take the divergence of equation (10) and employ Poisson's equation and the decomposition (2), we obtain the Newtonian version of Raychaudhuri's equation (Raychaudhuri 1955, 1957; Heckmann 1961):

$$\dot{\theta}+\tfrac{1}{3}\theta^2+2\sigma^2-2\omega^2+4\pi G\mu-\Lambda+\nabla^\alpha\left(\frac{1}{\mu}\nabla_\alpha p\right)=0. \tag{12}$$

This equation and the equation of continuity (9) are the key equations in this context; together with an equation of state of the form (1) they constitute a closed first-order system with respect to the variables μ, θ and p. We are seeking a second-order equation governing the density evolution, and proceed by taking the total time derivative of equation (9), $\dot{\mu}$ and $\dot{\theta}$ being eliminated using equations (9) and (12), to give

$$\frac{\ddot{\mu}}{\mu}-\frac{4}{3}\theta^2-2\sigma^2+2\omega^2-4\pi G\mu+\Lambda-\nabla^\alpha\left(\frac{1}{\mu}\nabla_\alpha p\right)=0. \tag{13}$$

This *exact* result is Hunter's equation (1964).

The variable introduced by Ellis & Bruni (1989) is

$$\mathscr{D}_\alpha=\frac{S\nabla_\alpha\mu}{\mu}=S\nabla_\alpha\int\frac{d\mu}{\mu} \tag{14}$$

where S is a local length-scale defined by

$$\frac{\dot{S}}{S}=\frac{1}{3}\theta. \tag{15}$$

The motivation behind this choice has been discussed at great length by Ellis & Bruni (1989), and will be mentioned here in the last section; for the moment we look for an exact second-order equation governing the evolution of $\mathscr{D}_\alpha$. We could proceed by taking the gradient of equation (13). The required result is obtained more easily, however, if we consider the gradients of the key first-order equations (9) and (12): we operate on equation (9) with ∇_α, and multiply the result by S; extra terms are introduced by commuting ∇_α and S with the total time derivative, for example

$$\begin{aligned} S\nabla_\alpha\dot{\mu}&=S(\nabla_\alpha\mu)^\cdot+S\mu_{;\gamma}\nabla_\alpha u^\gamma\\ &=(S\nabla_\alpha\mu)^\cdot+\left(\nabla_\alpha u^\gamma-\frac{1}{3}\theta\delta_\alpha{}^\gamma\right)S\nabla_\gamma\mu \\ &=(S\nabla_\alpha\mu)^\cdot+(\omega^\gamma{}_\alpha+\sigma^\gamma{}_\alpha)S\nabla_\gamma\mu \end{aligned} \tag{16}$$

using decomposition (2), to give

$$(S\nabla_\alpha\mu/\mu)^\cdot+(\omega^\beta{}_\alpha+\sigma^\beta{}_\alpha)S\nabla_\beta\mu/\mu+S\nabla_\alpha\theta=0 \tag{17}$$

i.e.

$$\dot{\mathscr{D}}_\alpha+(\omega^\beta{}_\alpha+\sigma^\beta{}_\alpha)\mathscr{D}_\beta+S\nabla_\alpha\theta=0. \tag{18}$$

In similar fashion, operation on equation (12) with ∇_α and multiplication of the result by S give

$$(S\nabla_\alpha\theta)^{\cdot}+(\omega^\beta{}_\alpha+\sigma^\beta{}_\alpha)S\nabla_\beta\theta$$
$$+S\nabla_\alpha\left[\frac{1}{3}\theta^2+2\sigma^2-2\omega^2+4\pi G\mu-\Lambda+\nabla^\beta\left(\frac{1}{\mu}\nabla_\beta p\right)\right]=0. \qquad (19)$$

Noting that $4\pi G\nabla_\alpha\mu=4\pi G\mu\mathscr{D}_\alpha$, we see that equations (18) and (19) together with an equation of state of the form (1) constitute a closed first-order system with respect to the variables $\mathscr{D}_\alpha$, $S\nabla_\alpha\theta$ and p, and the total time derivative of equation (18) leads to the desired second-order equation in $\mathscr{D}_\alpha$, which is

$$\ddot{\mathscr{D}}_\alpha+\frac{2}{3}\theta\dot{\mathscr{D}}_\alpha+2(\omega^\beta{}_\alpha+\sigma^\beta{}_\alpha)\dot{\mathscr{D}}_\beta$$
$$-4\pi G\mu\mathscr{D}_\alpha+(\omega^\beta{}_\alpha+\sigma^\beta{}_\alpha)^{\cdot}\mathscr{D}_\beta+(\omega^\beta{}_\alpha+\sigma^\beta{}_\alpha)(\omega^\gamma{}_\beta+\sigma^\gamma{}_\beta)\mathscr{D}_\gamma+\frac{2}{3}\theta(\omega^\beta{}_\alpha+\sigma^\beta{}_\alpha)\mathscr{D}_\beta$$
$$-2S\nabla_\alpha(\sigma^2-\omega^2)-S\nabla_\alpha\nabla^\beta\left(\frac{1}{\mu}\nabla_\beta p\right)=0. \qquad (20)$$

This *exact* equation is new.

4 RELATIVISTIC EQUATIONS

Key equations now are mass–energy conservation

$$\frac{\dot\mu}{\mu+p}+\theta=0 \qquad (21)$$

the equation of motion

$$\dot u_a+\frac{\nabla_a p}{\mu+p}=0 \qquad (22)$$

and Raychaudhuri's equation

$$\dot\theta+\frac{1}{3}\theta^2+2\sigma^2-2\omega^2+4\pi G(\mu+3p)-\Lambda-\dot u^a{}_{;a}=0. \qquad (23)$$

Equations (21) and (22) are $u^aT_a^b{}_{;b}=0$ and $h_a{}^bT_b{}^c{}_{;c}=0$ respectively. Raychaudhuri's equation is essentially $(u^a{}_{;ab}-u^a{}_{;ba})u^b=-R_{ab}u^au^b$, which arises from equation (6) if we contract on a and b and project along u^c. To make the parallel with Newtonian theory quite transparent, we must write the last term in Raychaudhuri's equation as

$$\dot u^a{}_{;a}=\dot u^a{}_{;b}\delta_a{}^b=\dot u^a{}_{;b}(h_a{}^b-u_au^b)=\nabla_a\dot u^a+\dot u_a\dot u^a$$
$$=-\nabla^a\left(\frac{1}{\mu+p}\nabla_a p\right)+\frac{\nabla^a p\nabla_a p}{(\mu+p)^2}. \qquad (24)$$

We now proceed by taking the total time derivative of equation (21), $\dot\mu$ and $\dot\theta$ being eliminated using equations (21) and (22), to give

$$\frac{\ddot\mu}{\mu+p}-\left(\frac{4}{3}+\frac{\mathrm{d}p}{\mathrm{d}\mu}\right)\theta^2-2\sigma^2+2\omega^2-4\pi G(\mu+3p)+\Lambda$$
$$-\nabla^a\left(\frac{1}{\mu+p}\nabla_a p\right)+\frac{\nabla^a p\nabla_a p}{(\mu+p)^2}=0. \qquad (25)$$

Equation (25) is the true relativistic generalization of Hunter's equation (13). A slightly more elegant formulation is achieved if we write $\nabla_a p$ as $p_{;a} + u_a \dot{p}$, to give

$$\frac{\ddot{\mu}}{\mu+p} - \frac{4}{3}\theta^2 - 2\sigma^2 + 2\omega^2 - 4\pi G(\mu+3p) + \Lambda - \nabla^a\left(\frac{1}{\mu+p} p_{;a}\right) + \dot{u}_a \dot{u}^a = 0 \tag{26}$$

which *exact* result is Jackson's equation (1972).

Turning now to the relativistic generalization of equation (20), we must first decide upon the appropriate analogue of the variable $\mathscr{D}_a$ defined by equation (14): Ellis & Bruni (1989) use $\mathscr{D}_a = S\nabla_a \mu/\mu$, but I find that a more natural choice is

$$\mathscr{D}_a = \frac{S\nabla_a \mu}{\mu+p} = S\nabla_a \int \frac{\mathrm{d}\mu}{\mu+p} \tag{27}$$

which choice results in considerable simplification. The Newtonian route now serves as a rough guide to further progress: we operate on equation (21) with ∇_a, and multiply the result by S; extra terms are introduced by commuting ∇_a and S with the total time derivative, for example

$$\begin{aligned} S\nabla_a \dot{\mu} &= S(\nabla_a \mu)^{\cdot} - S\dot{h}_a{}^b \mu_{;b} + S\mu_{;c} \nabla_a u^c \\ &= (S\nabla_a \mu)^{\cdot} - S\dot{u}_a \dot{\mu} - Su_a \dot{u}^b \mu_{;b} + \left(\nabla_a u^c - \frac{1}{3}\theta h_a{}^c\right) S\nabla_c \mu \\ &= (S\nabla_a \mu)^{\cdot} - S\dot{u}_a \dot{\mu} - Su_a \dot{u}^b \mu_{;b} + (\omega^c{}_a + \sigma^c{}_a) S\nabla_c \mu \end{aligned} \tag{28}$$

to give

$$h_a{}^b \dot{\mathscr{D}}_b + \left(\frac{\dot{p}}{\mu+p}\right) \mathscr{D}_a + (\omega^b{}_a + \sigma^b{}_a)\mathscr{D}_b + S\nabla_a \theta = 0. \tag{29}$$

Similarly, operation on equation (23) with ∇_a and multiplication by S gives

$$h_a{}^b (S\nabla_b \theta)^{\cdot} + \dot{\theta}\left(\frac{S\nabla_a p}{\mu+p}\right) + (\omega_a^b + \sigma_a^b) S\nabla_b \theta + S\nabla_a\left[\frac{1}{3}\theta^2 + 2\sigma^2 - 2\omega^2 + 4\pi G(\mu+3p) - \Lambda - \dot{u}^a{}_{;a}\right] = 0. \tag{30}$$

Noting that $4\pi G\nabla_a(\mu+3p) = 4\pi G(\mu+p)(1+3\mathrm{d}p/\mathrm{d}\mu)\mathscr{D}_a$, we see that equations (29) and (30) together with an equation of state of the form (1) constitute a closed first-order system with respect to the variables $\mathscr{D}_a$, $S\nabla_a \theta$ and p, and the total time derivative of equation (29) leads to the desired second-order equation in $\mathscr{D}_a$, which is

$$\begin{aligned} & h_a{}^b \ddot{\mathscr{D}}_b - \dot{u}_a \dot{u}^b \mathscr{D}_b + \left(\frac{\dot{p}}{\mu+p}\right)^{\cdot} \mathscr{D}_a + \left(\frac{\dot{p}}{\mu+p}\right) h_a{}^b \dot{\mathscr{D}}_b \\ & + h_a{}^b (\omega^c{}_b + \sigma^c{}_b)^{\cdot} \mathscr{D}_c + 2(\omega^b{}_a + \sigma^b{}_a) h_b{}^c \dot{\mathscr{D}}_c + (\omega^b{}_a + \sigma^b{}_a)(\omega^c{}_b + \sigma^c{}_b)\mathscr{D}_c \\ & + \left(\frac{\dot{p}}{\mu+p}\right)(\omega^b{}_a + \sigma^b{}_a)\mathscr{D}_b + \frac{2}{3}\theta h_a{}^b \dot{\mathscr{D}}_b + \frac{2}{3}\theta\left(\frac{\dot{p}}{\mu+p}\right)\mathscr{D}_a + \frac{2}{3}\theta(\omega^b{}_a + \sigma^b{}_a)\mathscr{D}_b \\ & - \dot{\theta}\left(\frac{S\nabla_a p}{\mu+p}\right) - 2S\nabla_a(\sigma^2 - \omega^2) - 4\pi G(\mu+p)\left(1 + 3\frac{\mathrm{d}p}{\mathrm{d}\mu}\right)\mathscr{D}_a \\ & + S\nabla_a(\dot{u}^c{}_{;c}) = 0. \end{aligned} \tag{31}$$

For purposes of checking, I have left the order of terms here as that in which they appear naturally in the transition from equations (29) and (30) to equation (31). Eliminating the terms in $\dot{u}^a$ using equations (22) and (24) and rearranging, we find

$$h_a^{\ b}\ddot{\mathscr{D}}_b+\left[\frac{2}{3}\theta+\left(\frac{\dot{p}}{\mu+p}\right)\right]h_a^{\ b}\dot{\mathscr{D}}_b+2(\omega^b_{\ a}+\sigma^b_{\ a})\,h_b^{\ c}\dot{\mathscr{D}}_c$$

$$+\left[\left(\frac{\dot{p}}{\mu+p}\right)^{\cdot}+\frac{2}{3}\theta\left(\frac{\dot{p}}{\mu+p}\right)-\dot{\theta}\frac{\mathrm{d}p}{\mathrm{d}\mu}-4\pi G(\mu+p)\left(1+3\frac{\mathrm{d}p}{\mathrm{d}\mu}\right)\right]\mathscr{D}_a$$

$$+h_a^{\ b}(\omega^c_{\ b}+\sigma^c_{\ b})^{\cdot}\mathscr{D}_c+(\omega^b_{\ a}+\sigma^b_{\ a})(\omega^c_{\ b}+\sigma^c_{\ b})\mathscr{D}_c$$

$$+\left[\frac{2}{3}\theta+\left(\frac{\dot{p}}{\mu+p}\right)\right](\omega^b_{\ a}+\sigma^b_{\ a})\mathscr{D}_b-\mathscr{P}_a\mathscr{P}^{\ b}\mathscr{D}_b-2S\nabla_a(\sigma^2-\omega^2) \tag{32}$$

$$+S\nabla_a(\mathscr{P}^{\ c}\mathscr{P}_c)-S\nabla_a\nabla^c\left(\frac{1}{\mu+p}\nabla_c p\right)=0$$

where

$$\mathscr{P}_a=\frac{\nabla_a p}{\mu+p}=S\nabla_a\int\frac{\mathrm{d}p}{\mu+p}. \tag{33}$$

Equation (32) is the relativistic generalization of equation (20).

5 THE PRESSURE TERM

To make further progress, we must consider the last terms in the propagation equations (20) and (32), which at the moment have the wrong form. With harmonic analysis in mind, the appropriate form in the context of the Newtonian equation (20) would be $S\nabla^\beta\nabla_\beta(\nabla_\alpha p/\mu)$. In this case, the commutation operation is trivial: as Galilean space is flat, we have $\nabla_\alpha\nabla^\beta=\nabla_\beta\nabla^\alpha$, and it is easy to show that $\nabla_\alpha(\mu^{-1}\nabla_\beta p)-\nabla_\beta(\mu^{-1}\nabla_\alpha p)=0$, for an equation of state of the form (1).

The corresponding undertaking in the context of the relativistic equation (32) is distinctly less trivial, owing to the appearance of the projection operator $h_a^{\ b}$ in the definition of ∇_a (equations 3 and 5), and to the non-commutativity of covariant differentiation in a non-flat space–time (equation 6). Nevertheless, it is not difficult to derive the following commutators (cf. Ellis et al. 1990):

$$\nabla_a\mathscr{P}_c-\nabla_c\mathscr{P}_a=\frac{-2\omega_{ac}\dot{p}}{\mu+p} \tag{34}$$

and

$$\nabla_a\nabla^c\mathscr{P}_c-\nabla^c\nabla_a\mathscr{P}_c=-2\omega_a^{\ c}h_c^{\ d}\dot{\mathscr{P}}_d-2K^c_{\ [a}K^d_{\ c]}\mathscr{P}_d+h_a^{\ d}h^{ef}R^c_{\ fed}\mathscr{P}_c \tag{35}$$

where $K_{ab}=h_a^{\ c}h_b^{\ d}u_{c;d}=\omega_{ab}+\sigma_{ab}+\frac{1}{3}\theta h_{ab}$.

At this stage we introduce the trace-free part of the Riemann tensor, the Weyl tensor, defined by

$$C^{ab}_{\ \ cd}=R^{ab}_{\ \ cd}-2\delta^{[a}_{[c}R^{b]}_{d]}+\tfrac{1}{3}R\delta^{[a}_{[c}\delta^{b]}_{d]}. \tag{36}$$

In conjunction with the field equations (7), equation (36) enables the last term in equation (35) to be written as

$$h_a^{\ d}h^{ef}R^c_{\ fed}\mathscr{P}_c=-\tfrac{2}{3}(\kappa\mu+\Lambda)\mathscr{P}_a-E^c_{\ a}\mathscr{P}_c \tag{37}$$

where $E_{cd}=u^a u^b C_{acbd}$ is the so-called electric part of the Weyl tensor. Combination of equation (34), (35) and (37) gives

$$\nabla_a\nabla^c\mathscr{P}_c-\nabla^c\nabla_c\mathscr{P}_a$$

$$=-\nabla^c\left(\frac{2\omega_{ac}\dot{p}}{\mu+p}\right)-2\omega_a^{\ c}h_c^{\ d}\dot{\mathscr{P}}_d-2K^c_{\ [a}K^d_{\ c]}\mathscr{P}_d-\tfrac{2}{3}(\kappa\mu+\Lambda)\mathscr{P}_a-E^c_{\ a}\mathscr{P}_c. \tag{38}$$

Ellis et al. (1989) originally based their approach to this term upon the concept of hypersurfaces approximately orthogonal to the fluid flow vector, and upon the equations of Gauss and Codazzi relating their curvature to that of the full space–time. As in general u^a is not hypersurface-orthogonal (i.e. $\omega_{ab}\neq 0$), their approach is perhaps a little artificial, and has indeed been the source of some confusion and error (see Ellis et al. 1990). I believe that my approach is more natural (and simpler), and that equation (38) is correct. Combination of equations (32) and (38) gives

$$h_a{}^b\ddot{\mathscr{D}}_b+\left[\frac{2}{3}\,\theta+\left(\frac{\dot p}{\mu+p}\right)\right]h_a{}^b\dot{\mathscr{D}}_b+2(\omega^b{}_a+\sigma^b{}_a)\,h_b{}^c\dot{\mathscr{D}}_c$$

$$+\left[\left(\frac{\dot p}{\mu+p}\right)^{\cdot}+\frac{2}{3}\theta\left(\frac{\dot p}{\mu+p}\right)-\dot\theta\frac{\mathrm{d}p}{\mathrm{d}\mu}-4\pi G(\mu+p)\left(1+3\frac{\mathrm{d}p}{\mathrm{d}\mu}\right)\right]\mathscr{D}_a$$

$$+h_a{}^b(\omega^c{}_b+\sigma^c{}_b)^{\cdot}\mathscr{D}_c+(\omega^b{}_a+\sigma^b{}_a)(\omega^c{}_b+\sigma^c{}_b)\,\mathscr{D}_c \tag{39}$$

$$+\left[\frac{2}{3}\,\theta+\left(\frac{\dot p}{\mu+p}\right)\right](\omega^b{}_a+\sigma^b{}_a)\,\mathscr{D}_b-\mathscr{P}_a\mathscr{P}^b\mathscr{D}_b-2S\nabla_a(\sigma^2-\omega^2)$$

$$+S\nabla_a(\mathscr{P}^c\mathscr{P}_c)-S\nabla^c\nabla_c\left(\frac{1}{\mu+p}\,\nabla_a p\right)+S\nabla^c\left(\frac{2\omega_{ac}\dot p}{\mu+p}\right)$$

$$+2\omega_a{}^c h_c{}^d S\dot{\mathscr{P}}_d+2K^c{}_{[a}K^d{}_{c]}S\mathscr{P}_d+\tfrac{2}{3}(\kappa\mu+\Lambda)\,S\mathscr{P}_a+E^c{}_aS\mathscr{P}_c=0.$$

Further simplifications are possible: for example, we can write

$$\frac{\dot p}{\mu+p}=\frac{\mathrm{d}p}{\mathrm{d}\mu}\,\frac{\dot\mu}{\mu+p}=-\frac{\mathrm{d}p}{\mathrm{d}\mu}\,\theta \tag{40}$$

using equation (21); the square-bracketed coefficient of $\mathscr{D}_a$ in equation (39) then simplifies using Raychaudhuri's equation (24), to give

$$h_a{}^b\ddot{\mathscr{D}}_b+\left[\frac{2}{3}-\left(\frac{\mathrm{d}p}{\mathrm{d}\mu}\right)\right]\theta h_a{}^d\dot{\mathscr{D}}_b+2(\omega^b{}_a+\sigma^b{}_a)\,h_b{}^c\dot{\mathscr{D}}_c$$

$$-\left[\left(\frac{\mathrm{d}p}{\mathrm{d}\mu}\right)^{\cdot}\theta+[4\pi G(\mu-3p)+2\Lambda-4(\sigma^2-\omega^2)-2\nabla^c\mathscr{P}_c+2\mathscr{P}^c\mathscr{P}_c]\frac{\mathrm{d}p}{\mathrm{d}\mu}\right]\mathscr{D}_a$$

$$-4\pi G(\mu+p)\mathscr{D}_a+h_a{}^b(\omega^c{}_b+\sigma^c{}_b)^{\cdot}\mathscr{D}_c+(\omega^b{}_a+\sigma^b{}_a)(\omega^c{}_b+\sigma^c{}_b)\mathscr{D}_c$$

$$+\left[\frac{2}{3}-\left(\frac{\mathrm{d}p}{\mathrm{d}\mu}\right)\right]\theta(\omega^b{}_a+\sigma^b{}_a)\mathscr{D}_b-\mathscr{P}_a\mathscr{P}^b\mathscr{D}_b-2S\nabla_a(\sigma^2-\omega^2)$$

$$+S\nabla_a(\mathscr{P}^c\mathscr{P}_c)-S\nabla^c\nabla_c\left(\frac{1}{\mu+p}\,\nabla_a p\right)-S\nabla^c\left(2\omega_{ac}\frac{\mathrm{d}p}{\mathrm{d}\mu}\,\theta\right)$$

$$+2\omega_a{}^c h_c{}^d S\dot{\mathscr{P}}_d+2K^c{}_{[a}K^d{}_{c]}S\mathscr{P}_d+\tfrac{2}{3}(\kappa\mu+\Lambda)\,S\mathscr{P}_a+E^c{}_aS\mathscr{P}_c=0. \tag{41}$$

This *exact* equation is new, and is the central result of this paper.

6 APPLICATION TO HOMOGENEOUS ISOTROPIC MODELS

In homogeneous isotropic models, there exist space-like hypersurfaces which are orthogonal to the flow vector u^a. These surfaces are taken to be surfaces of constant cosmic time t, on which for example the density μ is constant. A standard form of the corresponding metric is

$$\mathrm{d}s^2=-\mathrm{d}t^2+R(t)^2\left[\frac{\mathrm{d}r^2}{1-kr^2}+r^2(\mathrm{d}\theta^2+\sin^2\theta\,\mathrm{d}\phi^2)\right] \tag{42}$$

in which $k=\pm 1$, 0 is the appropriate curvature constant, and the field equations governing the evolution of the scale-factor $R(t)$ are

$$\ddot R=-\tfrac{4}{3}\pi G(\mu+3p)R+\tfrac{1}{3}\Lambda R \tag{43}$$

and

$$\dot{R}^2=\tfrac{8}{3}\pi G\mu R^2+\tfrac{1}{3}\Lambda R^2-k. \tag{44}$$

If we introduce another time coordinate,

$$t'=t-\epsilon(x) \tag{45}$$

where ϵ is an infinitesimal function of the spatial coordinates x, then surfaces of constant t' will not be homogeneous, and will contain density variations. This is the essence of the gauge problem: how do we distinguish between these fictitious variations and a genuinely inhomogeneous situation, with no space-like hypersurfaces that are exactly homogeneous and isotropic? In this context, $\mathscr{D}^a$ is a sensible choice as the dynamical variable: this vector field vanishes in the case of exact homogeneity and isotropy, so that any non-zero value cannot be entirely fictitious. This statement achieves precision if we introduce an orthonormal triad to vectors e^a_α, orthogonal also to u^a; we remove all degrees of freedom by demanding Fermi propagation along the flow lines:

$$\dot{e}^a_\alpha-(e^b_\alpha \dot{u}_b)\,u^a=0. \tag{46}$$

Such axes are analogous to a Newtonian non-rotating reference frame (see, for example, Trautman 1965). We could refer the propagation equation (39) to this frame, with $\mathscr{D}_\alpha=e^a_\alpha\mathscr{D}_a$, noting that

$$e^a_\alpha\dot{\mathscr{D}}_a=\dot{\mathscr{D}}_\alpha \tag{47}$$

and

$$e^a_\alpha\ddot{\mathscr{D}}_a=\ddot{\mathscr{D}}_\alpha+\mathscr{P}_\alpha\mathscr{P}^\beta\mathscr{D}_\beta \tag{48}$$

using equations (22) and (33). The vector $\mathscr{D}_a$ is proportional to the physical density gradient measured by a comoving observer, which cannot be a figment of his mathematical imagination.

Ellis & Bruni (1989) based their choice upon an argument outlined by Sachs (1984), and elaborated by Stewart & Walker (1974). In a classical perturbation approach, we would compare $\mathscr{D}_a$ at a point in one space–time with $\mathscr{D}_a$ in a nearby space–time, the two points being labelled by the same set of numbers. An infinitesimal coordinate transformation in, say, the second space–time shifts the point of comparison, and hence defines a vector field ϵ^a in said space–time. The corresponding change in $\mathscr{D}_a$ is equal to its Lie derivative with respect to this vector field, given by

$$\mathscr{L}_\epsilon\mathscr{D}^a=\mathscr{D}^a{}_{;c}\epsilon^c-\mathscr{D}^a\epsilon^a{}_{;c}. \tag{49}$$

Hence there is no ambiguity in the comparison if $\mathscr{L}_\epsilon\mathscr{D}^a$ vanishes for all vector fields ϵ^a, which is true if and only if $\mathscr{D}^a$ is exactly zero in one of the space–times, in our case the exactly homogeneous and isotropic one. Similarly, for a scalar ϕ we have $\mathscr{L}_\epsilon\phi=\phi_{,c}\epsilon^c$, so that a comparison is without ambiguity if and only if ϕ is a constant scalar.

We are now in a position to consider equation (39) when the space–time in question is approximately homogeneous and isotropic. When it is exactly so it is conformally flat, so that E_{ab} is exactly zero, as are the vorticity and shear tensors ω_{ab} and σ_{ab} and the dynamical variable $\mathscr{D}_a$ and all other spatial gradients; in the approximate case we expect to find coordinate patches in which these objects have components that are small, and a version of equation (41) in which products of such small quantities are neglected is

$$\begin{aligned} &h_a{}^b\ddot{\mathscr{D}}_b+\left(\frac{2}{3}-\frac{\mathrm{d}p}{\mathrm{d}\mu}\right)\theta h_a{}^b\dot{\mathscr{D}}_b \\ &\quad-\left[\left(\frac{\mathrm{d}p}{\mathrm{d}\mu}\right)^{\boldsymbol{\cdot}}\theta+[4\pi G(\mu-3p)+2\Lambda]\frac{\mathrm{d}p}{\mathrm{d}\mu}+4\pi G(\mu+p)\right]\mathscr{D}_a \\ &\quad-\nabla^c\nabla_c(S\mathscr{P}_a)-2\frac{\mathrm{d}p}{\mathrm{d}\mu}\theta S\nabla^c\omega_{ac}-\tfrac{2}{3}[\tfrac{1}{3}\theta^2-(\kappa\mu+\Lambda)]S\mathscr{P}_a=0. \end{aligned} \tag{50}$$

This equation is simpler than an equivalent one given by Ellis et al. (1990), which simplification arises from my definition of the key variable $\mathscr{D}_a$ (equation 27).

The unexpected feature of equations (1) and (50) is their lack of closure with respect to the variables $\mathscr{D}_a$ and $S\mathscr{P}_a$, occasioned by the first-order term in the vorticity gradient. The possibility of a corresponding vorticity-induced growth of the density gradient has been discussed by Ellis et al. (1990). However, the time evolution of vorticity is governed by the exact equation

$$h_a{}^c h_b{}^d\dot{\omega}_{cd}=-\omega_{ab}\left(\frac{2}{3}-\frac{\mathrm{d}p}{\mathrm{d}\mu}\right)\theta+2\sigma_{c[a}\omega_{b]}{}^c \tag{51}$$

(Hawking 1966); thus vorticity dies away rapidly as the Universe expands, for a reasonable equation of state. For this reason it seems unlikely that the density gradient growth induced by the extra term could be significant, other than at early times. In what follows, I shall assume that vorticity vanishes exactly; we note that equation (51) guarantees the persistence of such exactly irrotational motion.

The dividend arising from our investment in some intricate mathematics can now be claimed: to first order, equation (50) behaves as if the variables $\mathscr{D}_a$ and $S\mathscr{P}_a$ were decoupled from the space–time geometry, which is thus fixed by equation (42). The last term in equation (50) can also be simplified: in the FLRW case, the fluid flow vector u^a is exactly orthogonal to the space-like hypersurfaces of constant curvature, and the Ricci scalar $^*R$ induced on said hypersurfaces is given by

$$^*R=6\frac{k}{S^2}=2[-\tfrac{1}{3}\theta^2+\kappa\mu+\Lambda] \tag{52}$$

(see for example Ellis & Bruni 1989 and Hwang & Vishniac 1990), where $S(t)$ is now the scale-factor $R(t)$. Equation (52) is in fact the standard Friedmann equation (44). Thus, with $\omega_{ab}=0$, the last two terms in equation (50) become

$$-\nabla^2(S\mathscr{P}_a)+2(k/S^2)S\mathscr{P}_a=-\left(\frac{\mathrm{d}p}{\mathrm{d}\mu}\right)[\nabla^2\mathscr{D}_a-2(k/S^2)\mathscr{D}_a] \tag{53}$$

where ∇^2 is now the Laplacian on a three-space of constant curvature. Vector fields that are eigenfunctions of the Laplacian are discussed in many of the references mentioned here; in addition, lucid summaries have been given by Harrison (1967), Sandberg (1978) and Halliwell & Hawking (1985). Such fields are either divergence-free or curl-free; in view of equation (27), $\mathscr{D}_a$ is curl-free, and can be written as the divergence of a scalar harmonic Q^n, for which the eigenvalue structure is

$$\nabla^2 Q^n=-\frac{n^2-k}{S^2}Q^n \tag{54}$$

and

$$\nabla^2\nabla_a Q^n=-\frac{n^2-3k}{S^2}\nabla_a Q^n \tag{55}$$

where $n^2>0$ is continuous for $k=-1, 0$, and $n=2, 3, 4, \ldots$ is discrete for $k=+1$. Thus if

$$\mathscr{D}_a=\mathscr{D}^n(t)\nabla_a Q^n \tag{56}$$

then, according to equation (50), the amplitude $\mathscr{D}^n(t)$ satisfies

$$\begin{aligned}\ddot{\mathscr{D}}^n+&\left(\frac{2}{3}-\frac{\mathrm{d}p}{\mathrm{d}\mu}\right)\theta\dot{\mathscr{D}}^n\\ &-\left[\left(\frac{\mathrm{d}p}{\mathrm{d}\mu}\right)^{\bullet}\theta+[4\pi G(\mu-3p)-2\lambda]\frac{\mathrm{d}p}{\mathrm{d}\mu}+4\pi G(\mu+p)\right]\mathscr{D}^n\\ &+\left(\frac{\mathrm{d}p}{\mathrm{d}\mu}\right)\frac{n^2-k}{S^2}\mathscr{D}^n=0.\end{aligned} \tag{57}$$

In the simple case $p=w\mu$, where w is a constant, equation (57) becomes

$$\ddot{\mathscr{D}}^n+(\tfrac{2}{3}-w)\theta\dot{\mathscr{D}}^n-[4\pi G\mu(1-w)(1+3w)+2\Lambda w]\mathscr{D}^n+w\left(\frac{n^2-k}{S^2}\right)\mathscr{D}^n=0. \tag{58}$$

I consider now two particular applications of equation (58).

6.1 The Einstein static universe

In this model, with positive density μ and $k=+1$, the attractive effects of matter are cancelled by a positive cosmological constant

$$\Lambda = 4\pi G(\mu + 3p) = 4\pi G\mu(1 + 3w) \tag{59}$$

and the expansion θ vanishes. Equation (58) becomes

$$\ddot{\mathscr{D}}^n - [4\pi G\mu(1+w)(1+3w)]\,\mathscr{D}^n + w\left(\frac{n^2-1}{S^2}\right)\mathscr{D}^n = 0 \tag{60}$$

and the mode n with wavenumber k_n is unstable with respect to gravitational collapse if

$$4\pi G\mu(1+w)(1+3w) > w\left(\frac{n^2-1}{S^2}\right) = c_s^2 k_n^2 \tag{61}$$

which is the Jeans criterion in this setting; according to equation (52) it can be written as $1+3w > w(n^2-1)$, or $n^2 < 4 + 1/w$. The Einstein universe is unique in that it can be analysed without gauge problems using equation (25) (Jackson 1972); this is because μ and p are constant, so that according to the Sachs–Stewart–Walker theorem we can introduce unambiguous perturbations $\delta\mu$ and δp. With $\delta\mu/(\mu+p) = \mathscr{D}^n(t)\,Q^n$, the linearly perturbed version of equation (25) is exactly the same as equation (60), which is an important consistency check on this approach, and achieves the minor purpose mentioned in the second paragraph of the Introduction.

6.2 Spatially flat expanding universes, with $\Lambda = 0$ and $-1 \le w \le 1$

The range $|w| \le 1$ includes normal matter, $0 \le w \le 1/3$, and also corresponds to the effective equation of state associated with a homogeneous scalar field $\phi(t)$, the limit $\dot{\phi} \to \infty$ giving $w = +1$, i.e. $p - \mu = 0$, whereas $\dot{\phi} \to 0$ gives $w = -1$, i.e. $p + \mu = 0$. The latter case is the favoured model for a very early inflationary phase of the Universe; the possibility that such a field is dominant even at late times has also been taken seriously (see for example Ratra & Peebles 1988; Starkovich & Cooperstock 1992).

Equations (43) and (44) integrate to give

$$\mu \propto R^{-3(1+w)} \tag{62}$$

$$R \propto t^{2/3(1+w)^{-1}}. \tag{63}$$

I shall consider long-wavelength solutions only, so that the last term in equation (58) and the n-dependence can be ignored: $\mathscr{D}^n \to \mathscr{D}$; as $R(t)$ is a monotonically increasing function of time, we can regard $\mathscr{D}$ as a function of R, when equation (58) becomes

$$R^2\frac{d^2\mathscr{D}}{dR^2} + \tfrac{3}{2}(1-3w)R\frac{d\mathscr{D}}{dR} - \tfrac{3}{2}(1-w)(1+3w)\mathscr{D} = 0. \tag{64}$$

The two solutions to equation (64) are

$$\mathscr{D} \propto R^{(1+3w)} \tag{65}$$

and

$$\mathscr{D} \propto R^{-3(1-w)/2}. \tag{66}$$

Special cases are as follows:

(i) dust, $w=0$, giving a growing mode $\mathscr{D} \propto R$ and a decaying mode $\mathscr{D} \propto R^{-3/2}$;

(ii) pure radiation, $w = \frac{1}{3}$, giving a growing mode $\mathscr{D} \propto R^2$ and a decaying mode $\mathscr{D} \propto R^{-1}$;

(iii) stiff matter, $w=1$, giving a fixed mode and a growing mode $\mathscr{D} \propto R^4$, which is the maximum growth rate for $|w| \le 1$; it is tempting to invoke such matter, as a solution to whatever problems may arise with regard to the origin of significant inhomogeneities in the Universe, particularly as vorticity would also grow, according to equation (51); the problem here is that a phase dominated by such matter would almost certainly be short-lived, according to equation (62);

(iv) generalized inflation, $-\frac{1}{3} > w \ge -1 \Rightarrow \ddot{R} > 0$, for which there are only decaying modes; this is a generalized 'no hair' theorem for such models.

ACKNOWLEDGMENT

It is a pleasure to thank J. C. R. Magueijo of DAMTP Cambridge for a number of enlightening conversations about this topic.

REFERENCES

Bardeen J. M., 1980, Phys. Rev. D, 22, 1882
Bonnor W. B., 1957, MNRAS, 117, 104
Ellis G. F. R., 1990, MNRAS, 243, 509
Ellis G. F. R., Bruni M., 1989, Phys. Rev. D, 40, 1804
Ellis G. F. R., Hwang J., Bruni M., 1989, Phys. Rev. D, 40, 1819
Ellis G. F. R., Bruni M., Hwang J., 1990, Phys. Rev. D, 42, 1035
Halliwell J. J., Hawking S. W., 1985, Phys. Rev. D, 31, 1777
Harrison E. R., 1967, Rev. Mod. Phys., 39, 862
Hawking S. W., 1966, ApJ, 145, 544
Heckmann O., 1961, AJ, 66, 205
Heckmann O., Schücking E., 1955, Z. Astrophys., 38, 95
Heckmann O., Schücking E., 1956, Z. Astrophys., 40, 81
Hunter C., 1964, ApJ, 139, 570
Hwang J., Vishniac E. T., 1990, ApJ, 353, 1
Jackson J. C., 1972, Proc. R. Soc. Lond., A328, 561
Jeans J., 1902, Phil. Trans. R. Soc. Lond., A199, 49
Lifshitz E. M., 1946, J. Phys. USSR, 10, 116
Lifshitz E. M., Khalatnikov I. M., 1963, Adv. Phys., 12, 185
McCrea W. H., 1955, AJ, 60, 271
McCrea W. H., Milne E. A., 1934, Q. J. Math., 5, 73
Ratra B., Peebles P. J. E., 1988, Phys. Rev. D, 37, 3406
Raychaudhuri A., 1955, Phys. Rev., 98, 1123
Raychaudhuri A., 1957, Z. Astrophys., 43, 161
Sachs R. K., 1964, in DeWitt C., DeWitt B., eds, Les Houches Lectures 1963, Relativity, Groups and Topology. Blackie and Son, London, p. 556
Sandberg V. D., 1978, J. Math. Phys., 19, 2441
Starkovich S. P., Cooperstock F. I., 1992, ApJ, 389, 1
Stewart J. M., 1990, Class. Quantum Grav., 7, 1169
Stewart J. M., Walker M., 1974, Proc. R. Soc. Lond., A341, 49
Trautman A., 1965, in Brandeis Lectures 1964 Vol. 1, General Relativity. Prentice-Hall, New Jersey, pp. 137–142