International Journal of Theoretical Physics, Vol. 35, No. 3, 1996

Theory of Cosmological Perturbations Formulated in Terms of a Complete Set of Basic Gauge-Invariant Quantities

Zbigniew Banach[1] and Sławomir Piekarski[2]

Received June 6, 1995

An annoying paradox which has plagued the "naive" description of density perturbations in homogeneous and isotropic cosmological models has been the gauge-dependent character of this description. The corollary of this observation is that only gauge-invariant quantities have any inherent physical meaning. Thus the present paper develops, from a new geometric point of view, a totally gauge-invariant formulation of perturbation theory applicable to the case of a general perfect fluid with two essential thermodynamic variables. Precisely speaking, the main purpose here is the systematic construction of a complete set of basic gauge-invariant variables. This set consists of 17 linearly independent, not identically vanishing quantities. It turns out that these quantities can be used to divide the infinitesimal perturbations into equivalence classes: two perturbations P and P′ are said to be equivalent if their difference is equal to the Lie derivative of the background solution of Einstein's propagation equations with respect to an arbitrary vector field on the space-time manifold. In fact, the gauge-invariant perturbations, whose mathematical definition is best understood by introducing the elements of a certain quotient space, are uniquely determined from the basic variables. An additional welcome feature is that any gauge-invariant quantity can be constructed directly from the basic variables through purely algebraic and differential operations. In a companion paper, these results are used to derive the full, gauge-invariant system of equations governing the evolution of basic variables. In this sense, then, the present analysis is complete.

1. INTRODUCTION

The universe appears to be isotropic about every point in it and the mass distribution is close to homogeneous in the large-scale average, as far

[1] Centre of Mechanics, Institute of Fundamental Technological Research, Department of Fluid Mechanics, Polish Academy of Sciences, 00-049 Warsaw, Poland.
[2] Institute of Fundamental Technological Research, Department of Theory of Continuous Media, Polish Academy of Sciences, 00-049 Warsaw, Poland.

as we can detect. Thus the Robertson–Walker metrics or line elements are fundamental in the standard big-bang model of the universe. The mathematical framework in which the Robertson–Walker metrics occur is that of general relativity. Structure in the universe (galaxies, clusters of galaxies, etc.) is thought to have formed as a result of the growth, through gravitational collapse, of small spatial inhomogeneities in an otherwise smooth background ideal fluid. In order to understand the evolution of these small irregularities, one has to consider the equations of linear perturbation theory (Lifshitz, 1946; Lifshitz and Khalatnikov, 1963; Hawking, 1966) and over the past 15 years cosmologists have studied various solutions of such equations, usually only for the case of a barotropic perfect fluid where the pressure p and the energy density e are functionally dependent: $p = p(e)$.

The theory of linearized perturbations in its "naive" formulation has not been completely successful, however. Since the basic principles of general relativity are covariantly expressed, one easily verifies that any solution of the linearized field equations can be unique only up to a Lie derivative of the background solution with respect to an arbitrary vector field on the space-time manifold (Sachs, 1964; Ehlers, 1973). As a consequence, some cosmologists (Gerlach and Sengupta, 1978; Bardeen, 1980; Ellis and Bruni, 1989) are dissatisfied with the naive approach, for its equations appear to show that instead of a single solution to any particular set of gravitational circumstances there is an infinitude of equivalent solutions. The usual approach to the derivation of the deterministic equations governing linearized perturbations in cosmology has been to impose at the beginning a gauge condition to simplify the form of the metric and/or matter perturbations and then work directly with the metric tensor components and matter variables. In general, however, such approaches are plagued by the problem that the choice of variables to represent the density inhomogeneities in an almost-Robertson–Walker universe model depends on the gauge chosen. The corollary of these observations is that only gauge-invariant quantities can have any inherent physical meaning.

This is the first in a pair of articles (Banach and Piekarski, 1996), the overall objective of which is the systematic formulation of a new geometric approach to the theory of perturbations in homogeneous and isotropic cosmological models. Most of the discussion presented here will concern the case of a general perfect fluid where there are two essential thermodynamic variables (Misner *et al.*, 1973). We shall remove the necessity for referring to an infinitely large number of solutions to one problem at all by introducing a complete set of basic variables (17 linearly independent, not identically vanishing gauge-invariant quantities) that in no way depend on the choice of a vector field for the Lie derivative of the background. It turns out that these basic variables can be used to divide the infinitesimal perturbations into equivalence classes: two different solutions P and P′ of the linearized field equations are equivalent if there is a transformation of the Lie type which carries P into P′ and vice versa. Strictly speaking, then, the object of most physical interest is not just one perturbation P, but a whole equivalence class [P] of all perturbations P′ which are equivalent to P. In fact, there is something very special about the way the equivalence class [P] is related to the set of basic gauge-invariant variables; for once we completely specify this set, the equivalence class [P] is uniquely determined from it and conversely. An additional welcome feature is that any complicated gauge-invariant quantity can be constructed directly from the basic variables through purely algebraic and differential operations. Because of these two properties of the basic variables, our analysis seems to be complete.

In discussing the above problems, the starting point will be the development of a new geometric machinery associated with the existence of a preferred family of world lines and hence of a canonical scalar product in the space of gauge-dependent perturbations. The main idea we wish to pursue is that if this scalar product is systematically defined, one reasonable method for obtaining or generating a *countably* infinite set of gauge-invariant quantities is to use the standard procedure of orthogonalization (the Schmidt orthogonalization). Such a procedure can indeed be effectively applied, and some aspects of the original formulation of perturbation theory reconsidered. We shall study these issues quantitatively, at least in simple cases, in the first part of this article. More explicitly, based on the notion of a background solution, we define there a natural scalar product in the space of gauge-dependent perturbations by using the Robertson–Walker metric tensor and the unperturbed fluid four-velocity vector. Of course, given the gauge-invariant quantities, it is also useful to determine linear propagation equations for these quantities. We turn to this in the companion paper (Banach and Piekarski, 1996). Here it suffices to say that the resulting dynamical equations are deterministic, i.e., they lead to a unique solution for the basic gauge-invariant variables. Thus our approach sidesteps the usual problems. Nevertheless, it is perhaps important to stress that the gauge invariance in itself does not completely resolve the ambiguity of what one means by a physical solution to the linearized field equations, and some extra arguments are always necessary in order to eliminate the unphysical solutions of gauge-invariant equations. We discuss these new issues in Banach and Piekarski (1996, Sections 3.2 and 4.2). Here we only mention that such additional and unavoidable solutions of gauge-invariant equations *are not the same thing* as the spurious "gauge mode" solutions and thus *they must be distinguished from them.*

A totally gauge-invariant formulation of the linearized Einstein field equations was originally proposed by Gerlach and Sengupta (1978), and initial aspects of our formalism have been developed in various papers [see, e.g., Sachs (1964) and Ehlers (1973)]. Bardeen's (1980) major paper determined a set of gauge-invariant quantities that are related to density perturbations. Based on Hawking's pioneering analysis (Hawking, 1966), Ellis and Bruni (1989) gave in turn a simple alternative representation of density fluctuations. In the past, however, comparatively little attention has been focused upon the general problem of defining the equivalence class [P] in terms of a complete set of basic gauge-invariant variables; and, as far as we are aware, the explicit description of this equivalence class given here has not been obtained before. Moreover, much of the literature on relativistic hydrodynamics considers only barotropic perfect fluids. Our discussion will not make this "unphysical" restriction. Instead, the emphasis in this paper will be upon nonbarotropic perfect fluids where there are two essential thermodynamic variables, so that the simple equation of state $p = p(e)$ does not hold. This framework is sufficiently flexible and broadly based that it can be easily extended to materials more complex than we have considered here.

Another remark is also in order. Most analyses of inhomogeneities in an expanding universe have categorized the metric and matter perturbations into three distinct types: scalar, vector, and tensor perturbations. The generality of our geometric formulation allows us to show that the unique and transparent characterization of gauge-invariant perturbations is independent of this categorization. Nevertheless, in a companion paper (Banach and Piekarski, 1996) we briefly discuss how our gauge-invariant variables relate to those obtained by the technique of harmonic decomposition, which is also a very useful method in perturbation theory.

The program of this paper is as follows. In Section 2, we exploit the viewpoint that the direct way to formulate linear (or higher order) perturbation theory for the full nonlinear system of equations is to use one-parameter families of exact solutions to this system. In Section 3, some useful vector spaces are defined and the concept of the infinitesimal perturbation is introduced. In Sections 4 and 5, we define a complete set of basic gauge-invariant variables. In Section 6, we prove that this set enables the infinitesimal perturbations to be divided into equivalence classes: two different perturbations P and P′ are said to be equivalent if they differ by the action of an "infinitesimal diffeomorphism" on the background solution. Section 7 concludes the paper by summarizing its main results. In the Appendix, we show that any complicated gauge-invariant quantity is obtainable directly from the basic variables through purely algebraic and differential operations.

Our primary purpose here is to exhibit the general structure of perturbation theory. Thus attention will, in large part, focus upon such conceptual matters as how one might formulate exactly a notion of basic gauge-invariant variables in terms of which to characterize the time development of the perturbation. As regards the existence of linear perturbations, mathematical questions of this kind have been largely answered by D'Eath (1976). The systematic physical interpretation of gauge-invariant variables and the derivation of the linear propagation equations governing their evolution will be presented in a companion paper (Banach and Piekarski, 1996). Comparisons with other approaches are also given there.

2. PRELIMINARIES

2.1. Nonbarotropic Perfect Fluids

The fundamental equations of general relativity are Einstein's equations given by

$$R^{\alpha\beta} - \frac{1}{2}R^{\nu}{}_{\nu}\, g^{\alpha\beta} = T^{\alpha\beta} \tag{2.1}$$

where $R^{\alpha\beta}$ is the Ricci tensor; $R^{\nu}{}_{\nu}$ is the curvature tensor; $g^{\alpha\beta}$ is the contravariant metric tensor; and $T^{\alpha\beta}$ is the stress-energy tensor. We choose units so that the Einstein gravitational constant $8\pi G/c^4$ equals one ($8\pi G = c = 1$). Greek indices will range from 0 to 3, Latin indices from 1 to 3. It is assumed throughout that the space-time metric $g_{\alpha\beta}$ has signature $(-, +, +, +)$.

For a perfect fluid, $T^{\alpha\beta}$ takes the form

$$T^{\alpha\beta} = (e + p)u^{\alpha}u^{\beta} + pg^{\alpha\beta} \tag{2.2}$$

where e is the energy density, p is the pressure, and u^{α} is the four-velocity of matter ($u^{\alpha}u_{\alpha} = -1$). The values of e and p are assumed to be fixed uniquely by two thermodynamic variables, e.g., by the number density n and the temperature T as measured in the rest frame of the fluid element:

$$e = e(n, T), \qquad p = p(n, T) \tag{2.3}$$

In this sense the fluid is a nonbarotropic perfect fluid. The number flux vector, namely

$$N^{\alpha} = nu^{a} \tag{2.4}$$

obeys the "continuity law"

$$N^{\alpha}{}_{;\alpha} = 0 \tag{2.5}$$

Here, of course, a semicolon denotes the covariant derivative of N^{α} with respect to $g_{\alpha\beta}$.

With equations (2.3), the continuity law and Einstein's field equations place five constraints (balance of number density and local conservation of energy and momentum) on N^{α} and $T^{\alpha\beta}$ and six constraints on the ten $g_{\alpha\beta}$, leaving four of the $g_{\alpha\beta}$ to be adjusted by the choice of the "gauge condition."

2.2. One-Parameter Families of Exact Solutions

In an exact description, the full nonlinear system of equations, which consists of equations (2.1)–(2.5), would become a complicated set of equations for the evaluation of $\mathcal{G} := (g^{\alpha\beta}, u^{\alpha}, n, T)$. Solving this system is not simple, but it turns out that if the matter is only slightly perturbed away from the background Robertson–Walker cosmological model, then another device worth noting is that of using perturbation theory to obtain the linearized system of field equations.

The basic assumptions of this theory, which seem necessary in order to give a clear idea of what the perturbation method is to be (Ehlers, 1973; Banach and Piekarski, 1994a), may be formulated as follows: Consider an open interval $I := (-d, d)$ of R, $d > 0$. For each $\epsilon \in I$ there exists a classical solution $\mathcal{G}_\epsilon$ to the full nonlinear system of equations

$$\mathcal{G}_\epsilon := (g^{\alpha\beta}(\epsilon,\cdot),\ u^{\alpha}(\epsilon,\cdot),\ n(\epsilon,\cdot),\ T(\epsilon,\cdot)) \tag{2.6}$$

such that the objects appearing on the right-hand side of equation (2.6) are continuously differentiable with respect to $\epsilon \in I$. The set of classical solutions defines a function space, and a one-parameter family of exact solutions given by $\{\mathcal{G}_\epsilon;\ \epsilon \in I\}$ may be thought of as a curve in the function space passing through the "point" $\mathcal{G}_0$ which we call the background solution. [As already remarked by Bardeen (1980) and Ellis and Bruni (1989), in a cosmological setting mathematical questions regarding the validity of perturbation theory and the existence of such curves have been largely answered by D'Eath (1976). The objections raised to this method by Fischer *et al.* (1980) do not apply here, because the solutions $\mathcal{G}_\epsilon$ contain matter; see also the discussion in Section 2.3 of Banach and Piekarski (1996).]

The central theme of perturbation theory is the construction of the mapping $\epsilon \Rightarrow \mathcal{G}_\epsilon$ for small values of ϵ, starting with a given solution $\mathcal{G}_0$, which may be the Robertson–Walker solution. The "tangent" to the curve $\epsilon \Rightarrow \mathcal{G}_\epsilon$ considered at $\epsilon = 0$ is defined by

$$\delta\mathcal{G}_0 := (Q^{\alpha\beta},\ U^{\alpha},\ n_0 M,\ T_0 K) \tag{2.7}$$

where

$$Q^{\alpha\beta} := \left(\frac{\partial g^{\alpha\beta}}{\partial \epsilon}\right)_{|\epsilon=0}, \qquad U^{\alpha} := \left(\frac{\partial u^{\alpha}}{\partial \epsilon}\right)_{|\epsilon=0} \tag{2.8a}$$

$$n_0 := (n)_{|\epsilon=0}, \qquad T_0 := (T)_{|\epsilon=0}, \qquad T_0 > 0 \tag{2.8b}$$

$$M := \frac{1}{n_0}\left(\frac{\partial n}{\partial \epsilon}\right)_{|\epsilon=0}, \qquad K := \frac{1}{T_0}\left(\frac{\partial T}{\partial \epsilon}\right)_{|\epsilon=0} \tag{2.8c}$$

Clearly, the pair $(\mathcal{G}_0, \delta\mathcal{G}_0)$ "models" the curve $\epsilon \Rightarrow \mathcal{G}_\epsilon$ at $\epsilon = 0$. Just as in Ehlers (1973), the tangent $\delta\mathcal{G}_0$ is said to be an *infinitesimal* perturbation of $\mathcal{G}_0$. Interpreting n_0 and T_0, we call n_0 the background number density and T_0 the background temperature. We gain an intuitive feeling for the meaning of M and K if we say that they are, respectively, the fractional variations in density and temperature along a world line. The quantities $Q^{\alpha\beta}$ and U^α represent in turn changes in the metric and in the matter velocity.

To derive a closed set of governing equations for perturbations for the system, we must differentiate Einstein's field equations (2.1) and the equation of balance of number density (2.5) with respect to ϵ and then set ϵ equal to zero:

$$\left[\frac{\partial}{\partial\epsilon}\left(R^{\alpha\beta} - \frac{1}{2}R^\nu{}_\nu g^{\alpha\beta}\right)\right]_{|\epsilon=0} = \left(\frac{\partial T^{\alpha\beta}}{\partial\epsilon}\right)_{|\epsilon=0} \tag{2.9a}$$

$$\left(\frac{\partial}{\partial\epsilon}N^\alpha{}_{;\alpha}\right)_{|\epsilon=0} = 0 \tag{2.9b}$$

Equations (2.9a) and (2.9b) are linear equations for $\delta\mathcal{G}_0$, i.e., they can be expressed in the form (Wald, 1984)

$$\mathcal{E}(\delta\mathcal{G}_0) = 0 \tag{2.10}$$

where $\mathcal{E}$ is a linear differential space-time operator acting on $\delta\mathcal{G}_0$. If we can solve equation (2.10) for $\delta\mathcal{G}_0$, then $\mathcal{G}_0 + \epsilon\,\delta\mathcal{G}_0$ should yield a good approximation to $\mathcal{G}_\epsilon$ for sufficiently small ϵ, and issues of cosmological interest thus can be investigated.

As discussed in detail by Ehlers (1973) and Wald (1984), there is a gauge freedom in general relativity corresponding to the group of diffeomorphisms of space-time. Within the framework of a linear approximation, this implies that two perturbations $\delta\mathcal{G}_0$ and $\delta\mathcal{G}_0'$ represent the same perturbation if (and only if) they differ by the action of an "infinitesimal diffeomorphism" on the background solution $\mathcal{G}_0$. An infinitesimal diffeomorphism and its action on $\mathcal{G}_0$ can be described in terms of a vector field v on the space-time manifold X. Precisely speaking, elementary inspection shows that the change in a perturbation induced by v is the Lie derivative $\mathcal{L}_v$ of $\mathcal{G}_0$ with respect to v. Thus $\delta\mathcal{G}_0$ and $\delta\mathcal{G}_0 + \mathcal{L}_v\mathcal{G}_0$ represent the same physical perturbation (the equivalence class $[\delta\mathcal{G}_0]$ of $\delta\mathcal{G}_0$), and clearly $\delta\mathcal{G}_0$ satisfies the linearized field equations (2.10) if and only if $\delta\mathcal{G}_0 + \mathcal{L}_v\mathcal{G}_0$ does.

The usual approach to the derivation of the equations governing linearized perturbations in cosmology has been to impose at the beginning a gauge condition to simplify the form of the tangent $\delta\mathcal{G}_0$. However, as observed already by Ellis and Bruni (1989) [see also Bardeen's (1980) major paper], "the resulting problem is that the quantity $n_0 M$ (the variation in density along a single world line) often calculated in perturbation calculations is completely dependent on the gauge chosen, and unless this gauge is fully specified the modes discovered for this quantity are spurious modes (due to residual gauge freedom); while if it is fully specified, its relation to what we really want to know (the spatial variation of density in the universe) is convoluted and difficult to interpret."

The corollary of this observation is that only a gauge-invariant approach to cosmological density fluctuations can have any inherent physical meaning. Consequently, the object here and in a companion paper (Banach and Piekarski, 1996) will be the systematic development of such an approach.

3. FURTHER DEFINITIONS CONCERNING THE INFINITESIMAL PERTURBATION

3.1. Some Useful Vector Spaces

At this stage, one is in a position to address the issue of deriving the concept of an infinitesimal perturbation from a slightly more systematic point of view. However, before passing on to this problem, it is important to first define some useful vector spaces. This will, among other things, serve to clarify the geometric content of perturbation theory described in Section 2.2.

Consider a local coordinate system (x^α) with four functions x^α, $\alpha = 0, \ldots, 3$, whose values $x^\alpha(x)$ are the coordinates of the point x of the space-time manifold X. If we define b_α and $b_{\alpha\beta}$ by

$$b_\alpha := \frac{\partial}{\partial x^\alpha} \tag{3.1a}$$

and

$$b_{\alpha\beta} := \frac{1}{2}\left(\frac{\partial}{\partial x^\alpha} \otimes \frac{\partial}{\partial x^\beta} + \frac{\partial}{\partial x^\beta} \otimes \frac{\partial}{\partial x^\alpha}\right) \tag{3.1b}$$

then $\{b_\alpha\}$ is a coordinate basis for vectors and $\{b_{\alpha\beta}\}$ is a coordinate basis for all symmetric contravariant tensors of rank two. Now, using the terminology and notation of Section 2.2 [see, e.g., equation (2.6)], it follows that for each $\epsilon \in I$, $g^{\alpha\beta}(\epsilon,x)$ are the components of the contravariant metric tensor $g(\epsilon,x)$ with respect to $\{b_{\alpha\beta}\}$:

$$g(\epsilon,x) = g^{\alpha\beta}(\epsilon,x) b_{\alpha\beta} \tag{3.2}$$

Similarly, $u^\alpha(\epsilon,x)$ are the components of the normalized four-velocity $u(\epsilon,x)$ with respect to $\{b_\alpha\}$:

$$u(\epsilon,x) = u^\alpha(\epsilon,x) b_\alpha \tag{3.3}$$

For each $x \in X$ in a coordinate domain, let V_x denote the set of all symmetric contravariant tensors of rank two and let W_x denote the set of all contravariant vectors; these tensors and vectors are defined at x. It is obvious that V_x and W_x are vector spaces. Moreover, given V_x and W_x, one easily verifies that

$$g(\epsilon,x) \in V_x, \qquad u(\epsilon,x) \in W_x \tag{3.4}$$

This conclusion holds for each $\epsilon \in I$, even though $g(\epsilon,x)$ is a nonsingular second-rank, symmetric tensor and $u(\epsilon,x)$ is a timelike, unit-magnitude vector:

$$\det\|g^{\alpha\beta}(\epsilon,x)\| \neq 0 \tag{3.5a}$$

$$g_{\alpha\beta}(\epsilon,x)u^{\alpha}(\epsilon,x)u^{\beta}(\epsilon,x) = -1 \tag{3.5b}$$

From equation (3.5b) it follows at once that the four-velocity $u^{\alpha}(\epsilon,\cdot)$ cannot be defined without first introducing the metric $g_{\alpha\beta}(\epsilon,\cdot)$.

In this paper, the background space-time considered is described by a Robertson–Walker metric. We shall present the detailed calculations only for the case of zero spatial curvature, $k = 0$, in order to simplify the discussion. However, our approach can easily be extended to allow for nonzero background three-space curvature ($k = \pm 1$), and we will deal with the most important aspects of this extension in Sections 5.2 and 6. By an appropriate choice of coordinates for the space-time manifold X, the contravariant metric tensor

$$g_{(0)}(x) = g_{(0)x} := g(\epsilon,x)_{|\epsilon=0} \tag{3.6}$$

then reduces to the familiar form

$$g_{(0)}(x) = -b_{00} + R^{-2}(t)\delta^{rs}b_{rs} \tag{3.7}$$

where $t := x^0$ and where $R(t)$ is the expansion factor. As to the meaning of δ^{rs}, this symbol represents the Kronecker delta. Of course, in equation (3.7) Einstein's summation convention is used.

With these notions, the geometrically preferred four-velocity

$$u_{(0)}(x) := u(\epsilon,x)_{|\epsilon=0} \tag{3.8}$$

can be written as

$$u_{(0)}(x) = b_0 \tag{3.9}$$

Let $\{dx^{\alpha}\}$ be a basis of one-forms dual to a basis $\{b_{\alpha}\}$. Further, introduce the following useful abbreviations:

$$b^{\alpha} := dx^{\alpha} \tag{3.10a}$$

$$b^{\alpha\beta} := \frac{1}{2}(dx^{\alpha} \otimes dx^{\beta} + dx^{\beta} \otimes dx^{\alpha}) \tag{3.10b}$$

Denoting by g_x^{RW} the $k = 0$ Robertson–Walker metric, we then find from equation (3.7) that

$$g_x^{\mathrm{RW}} = -b^{00} + R^2(t)\delta_{rs}b^{rs} \tag{3.11}$$

where again δ_{rs} represents the Kronecker delta.

It will be convenient to introduce the quantity $\bar{n}(\epsilon,x)$, which measures the relative size of the "actual" number density $n(\epsilon,x)$ compared to the background number density $n_0(t)$. Thus we set

$$\bar{n}(\epsilon,x) := \frac{1}{n_0(t)}\, n(\epsilon,x) \in \mathrm{R}_1 := \mathrm{R} \tag{3.12a}$$

Accordingly, we define $\bar{T}(\epsilon,x)$ by

$$\bar{T}(\epsilon,x) := \frac{1}{T_0(t)}\, T(\epsilon,x) \in \mathrm{R}_2 := \mathrm{R} \tag{3.12b}$$

where $T(\epsilon,x)$ is the "actual" temperature and $T_0(t)$ is the background temperature. Here R_1 and R_2 are one-dimensional vector spaces, each endowed with the canonical structure of a set of real numbers: $\mathrm{R}_1 = \mathrm{R}_2 = \mathrm{R}$. In view of equations (3.12), we regard $\bar{n}(\epsilon,x)$ as a member of R_1 and $\bar{T}(\epsilon,x)$ as a member of R_2.

Some aspects of perturbation theory can be most easily discussed by constructing a vector space $\mathcal{W}_x$, which is the (external) direct sum of V_x, W_x, R_1, and R_2 (Greub, 1975). The manner in which

$$\mathcal{W}_x := V_x \oplus W_x \oplus \mathrm{R}_1 \oplus \mathrm{R}_2 \tag{3.13}$$

and

$$\mathcal{W} := \bigcup_{x\in X} \mathcal{W}_x \tag{3.14}$$

form the natural geometric objects defined on the space-time manifold X will become clear in the text below.

3.2. The Structure of $\mathcal{W}$

One should keep in mind that $\mathcal{W}$ is the vector bundle over a space-time manifold X obtained by giving $\mathcal{W} := \bigcup_{x\in X} \mathcal{W}_x$ its natural structure and its natural projection onto X (Choquet-Bruhat *et al.*, 1989). Thus the projection π maps each point of $\mathcal{W}_x$ into x. The idea underlying the introduction of $\mathcal{W}$ is simply this: After specifying $\mathcal{G}_0$ and replacing

$$\mathcal{G}_\epsilon(x) := (g^{\alpha\beta}(\epsilon,x), u^\alpha(\epsilon,x), n(\epsilon,x), T(\epsilon,x)) \tag{3.15}$$

by

$$C(\epsilon,x) := g(\epsilon,x) \oplus u(\epsilon,x) \oplus \bar{n}(\epsilon,x) \oplus \bar{T}(\epsilon,x) \tag{3.16}$$

as is always possible, the solution $\mathcal{G}_\epsilon$ of Einstein's field equations and the equation of balance of number density can be identified with an appropriate cross section of $\mathcal{W}$; this cross section is given by $x \Rightarrow \mathcal{G}_\epsilon(x)$ or by

$$X \ni x \Rightarrow C(\epsilon,x) \in \mathcal{W}_x \tag{3.17}$$

The "smooth" curve $\epsilon \Rightarrow \mathcal{G}_\epsilon$ discussed in Section 2.2, which consists of exact solutions to the aforementioned system of field equations, is thus a one-parameter family of suitably chosen cross sections of the vector bundle $\mathcal{W}$. In the rest of this paper, we shall always make such an identification.

Put

$$C_0(x) := C(\epsilon,x)_{|\epsilon=0} = g_{(0)}(x) \oplus u_{(0)}(x) \oplus 1 \oplus 1 \tag{3.18}$$

so that the mapping $x \Rightarrow C_0(x)$ is a "representation" of the background solution. An infinitesimal perturbation considered at $x \in X$ is by definition a pair $(C_0(x),\mathrm{P}(x))$ in which $C_0(x)$ and $\mathrm{P}(x)$ are given by equation (3.18) and

$$\mathrm{P}(x) := \left(\frac{\partial C(\epsilon,x)}{\partial \epsilon}\right)_{|\epsilon=0} \tag{3.19}$$

respectively. In view of equations (2.8) and (3.12), we conclude that

$$\mathrm{P}(x) = Q(x) \oplus U(x) \oplus M(x) \oplus K(x) \tag{3.20}$$

where

$$Q(x) := Q^{\alpha\beta}(x)b_{\alpha\beta} \in V_x \tag{3.21a}$$

and

$$U(x) := U^{\alpha}(x)b_{\alpha} \in W_x \tag{3.21b}$$

Further, from equations (2.8c) and (3.12) it follows that $M(x) \in \mathrm{R}_1$ and $K(x) \in \mathrm{R}_2$. Consequently, $\mathrm{P}(x)$ is an element of $\mathcal{W}_x$. By abuse of language, instead of saying that (C_0,P) is an infinitesimal perturbation, we shall also say that P itself is an infinitesimal perturbation.

Clearly, using the background solution $\mathcal{G}_0$ and the mapping $x \Rightarrow \mathrm{P}(x)$, one can easily calculate the tangent $\delta\mathcal{G}_0$ satisfying equation (2.10). In this way of thinking, the mapping

$$X \ni x \Rightarrow \mathrm{P}(x) \in \mathcal{W}_x \tag{3.22}$$

can be interpreted as a solution to the linearized system of field equations.

Another remark is also in order. In the discussion of direct sums of families of vector spaces [see Greub (1975, p. 56)] it will be convenient not to distinguish between V_x, W_x, R_1, and R_2 and their images in $\mathcal{W}_x$ under the canonical injections, but to regard them as the same vector spaces. Thus, for notational convenience, the images in $\mathcal{W}_x$ of V_x, W_x, R_1, R_2 are simply denoted by V_x, W_x, R_1, R_2, respectively. Obviously, because of this convention, we are justified in saying that the contravariant metric tensor $\underline{g}(\epsilon,x)$ is an element of $\mathcal{W}_x$. The same conclusion holds for $u(\epsilon,x)$, $\bar{n}(\epsilon,x)$, and $\overline{T}(\epsilon,x)$. However, it will always be clear from the context whether, e.g., the interpretation $g(\epsilon,x) \in V_x$ or $g(\epsilon,x) \in \mathcal{W}_x$ is meant.

3.3. The Lie Derivative

We have already remarked in Section 2.2 that if

$$v = v^\alpha b_\alpha \tag{3.23}$$

is an arbitrary vector field on X, then $\delta\mathcal{G}_0$ and $\delta\mathcal{G}_0 + \mathcal{L}_v\mathcal{G}_0$ represent the same physical perturbation. Here, of course, $\mathcal{L}_v\mathcal{G}_0$ gives rise to the notion of the Lie derivative $\mathcal{L}_v$ with respect to v. To analyze the action of $\mathcal{L}_v$ on the background tensor fields, it is helpful and natural to choose a coordinate system (x^α) on X so that the unperturbed objects $g_{(0)}$ and $u_{(0)}$ are defined by equations (3.7) and (3.9) and the background scalar quantities n_0, T_0 are functions of the time coordinate $t := x^0$ only. (This always can be done locally in any region of space-time.) Thus we obtain for $\mathcal{L}_v g_{(0)}$, $\mathcal{L}_v u_{(0)}$, $\mathcal{L}_v n_0$, and $\mathcal{L}_v T_0$ the following formulas (Choquet-Bruhat *et al.*, 1989):

$$\begin{aligned}\mathcal{L}_v g_{(0)} = {} & 2v^0_{,0}b_{00} + 2(-R^{-2}\delta^{rp}v^0_{,p} + v^r_{,0})b_{0r} \\ & - [2R^{-2}v^0H\delta^{rs} + R^{-2}(\delta^{rp}v^s_{,p} + \delta^{sp}v^r_{,p})]b_{rs}\end{aligned} \tag{3.24a}$$

$$\mathcal{L}_v u_{(0)} = -v^\alpha_{,0}b_\alpha \tag{3.24b}$$

$$\mathcal{L}_v n_0 = v^0\dot{n}_0, \qquad \mathcal{L}_v T_0 = v^0\dot{T}_0 \tag{3.24c}$$

where a comma denotes the partial derivative in X and an overdot indicates differentiation with respect to time. As to the meaning of H, this is Hubble's parameter given by

$$H := \dot{R}/R \tag{3.25}$$

Turning our attention back to equation (3.18) and its interpretation, what we must do now is to specify explicitly the action of $\mathcal{L}_v$ on C_0. After a bit of mathematical manipulation which employs only the obvious definition

$$\mathcal{L}_v\mathcal{G}_0 := ((\mathcal{L}_v g_{(0)})^{\alpha\beta}, (\mathcal{L}_v u_{(0)})^\alpha, \mathcal{L}_v n_0, \mathcal{L}_v T_0) \tag{3.26}$$

we find from equation (3.16) that the "best" expression for $\mathcal{L}_v C_0$, in the sense that it comes nearest to $\mathcal{L}_v \mathcal{G}_0$, is given by

$$\mathcal{L}_v C_0 := \mathcal{L}_v g_{(0)} \oplus \mathcal{L}_v u_{(0)} \oplus \overline{\mathcal{L}_v n_0} \oplus \overline{\mathcal{L}_v T_0} \tag{3.27}$$

where

$$\overline{\mathcal{L}_v n_0} := \frac{1}{n_0} \mathcal{L}_v n_0 = \frac{1}{n_0} v^0 \dot{n}_0 = -3v^0 H \tag{3.28a}$$

$$\overline{\mathcal{L}_v T_0} := \frac{1}{T_0} \mathcal{L}_v T_0 = \frac{1}{T_0} v^0 \dot{T}_0 \tag{3.28b}$$

In obtaining equation (3.28a), we have made use of the fact that $\dot{n}_0 = -3n_0 H$ (Peebles, 1993). Since the operation $\mathcal{L}_v$A on a tensor A produces a tensor of the same rank as A, the mapping

$$X \ni x \Rightarrow (\mathcal{L}_v C_0)(x) \in \mathcal{W}_x \tag{3.29}$$

can be viewed as a cross section of $\mathcal{W}$.

The set consisting of the mappings (3.29) for all differentiable vector fields v on X is written $\mathcal{P}_0$; this set carries a natural structure of a vector space. Clearly, $\mathcal{P}_0$ is a subspace of the space $\mathcal{P}$ *whose elements are classical solutions to the linearized system of field equations*. Two infinitesimal perturbations P $\in \mathcal{P}$ and P$'$ $\in \mathcal{P}$ will be taken to be equivalent if there is a vector field v on X such that P$'$ = P + $\mathcal{L}_v C_0$. Strictly speaking, then, the object of most physical interest is not just one perturbation P, but a whole equivalence class of all perturbations P$'$ which are equivalent to P. This equivalence class is denoted [P] and is called the *gauge-invariant perturbation* associated with P. Thus the gauge-invariant perturbations are elements of $\mathcal{P}/\mathcal{P}_0$, the quotient space of $\mathcal{P}$ by $\mathcal{P}_0$ (Choquet-Bruhat *et al.*, 1989). The essential point in the theory of gauge-invariant perturbations is to describe the elements of this quotient space explicitly. These issues will be discussed in Section 6.

Finally, we pass to the problem of calculating $\mathcal{L}_v C_0$. By combining equations (3.24) and (3.27), it is only a matter of labor to prove that

$$\begin{aligned}(\mathcal{L}_v C_0)(x) = {} & -3v^0 H\left[\left(\frac{2}{3} R^{-2}\delta^{rs} b_{rs}\right) \oplus 1 \oplus (-(3HT_0)^{-1}\dot{T}_0)\right] \\ & - v^0_{,0}(-2b_{00} \oplus b_0) - (\delta^{rp} v^s_{,p} + \delta^{sp} v^r_{,p})(R^{-2} b_{rs}) \\ & - 2(R^{-1}\delta^{rp} v^0_{,p} - R v^r_{,0})(R^{-1} b_{0r}) - R v^r_{,0}(R^{-1} b_r)\end{aligned} \tag{3.30}$$

where

$$b_{\alpha\beta} \in V_x \subset \mathcal{W}_x, \qquad b_\alpha \in W_x \subset \mathcal{W}_x \tag{3.31a}$$

$$1 \in \mathrm{R}_1 \subset \mathcal{W}_x, \qquad -(3HT_0)^{-1}\dot{T}_0 \in \mathrm{R}_2 \subset \mathcal{W}_x \tag{3.31b}$$

Because of the identification rule of Section 3.2, in equation (3.30) no distinction is made between the *external* and *internal* direct sums (Greub, 1975). Moreover, the dependence of n_0, T_0, R, H, and v^α on t or x is not shown explicitly, in order to make the resulting formulas shorter. What equation (3.30) really entails, and why it should be important, will be discussed more carefully in Section 4.1.

3.4. The Scalar Product in $\mathcal{W}_x$

To carry on the intended analysis of the properties of gauge-invariant perturbations, it is useful to define a scalar product in $\mathcal{W}_x$. Based on the notion of a background solution, we arrive in fact at the canonical definition by using the covariant image of $g(\epsilon,x)_{|\epsilon=0}$ (which is the $k = 0$ Robertson–Walker metric tensor g_x^{RW}) and the unperturbed four-velocity vector (which is given by $u_{(0)}(x) = b_0$). Thus, for an almost-Robertson–Walker universe model, the symmetry is broken by the existence of a preferred vector field $u_{(0)}$. This "phenomenon" of symmetry breaking plays an important role in the present formulation of perturbation theory, because it enables us to understand the meaning of the following construction:

1. Let $W_{x,1}$ be a linear closure of $\{u_{(0)}(x)\}$, and denote by $W_{x,2}$ the g_x^{RW}-orthogonal complement of $W_{x,1}$. Clearly, $W_{x,1}$ and $W_{x,2}$ are vector subspaces of W_x; moreover, we have

$$W_x = W_{x,1} \oplus W_{x,2} \tag{3.32}$$

2. Next, define the vector subspaces $V_{x,1}$, $V_{x,2}$, and $V_{x,3}$ of V_x by setting

$$V_{x,1} := W_{x,1} \otimes W_{x,1} \tag{3.33a}$$

$$V_{x,2} := \Pi(W_{x,1} \otimes W_{x,2}) \tag{3.33b}$$

$$V_{x,3} := \Pi(W_{x,2} \otimes W_{x,2}) \tag{3.33c}$$

Here $V_{x,2}$ is the image space of the symmetrizer Π in $W_{x,1} \otimes W_{x,2}$ and $V_{x,3}$ is the image space of the symmetrizer Π in $W_{x,2} \otimes W_{x,2}$ (Meyer and Schröter, 1981; Banach and Piekarski, 1989). Given the definitions (3.33), we can easily verify that

$$V_x = V_{x,1} \oplus V_{x,2} \oplus V_{x,3} \tag{3.34}$$

is a decomposition of V_x as a direct sum of $V_{x,1}$, $V_{x,2}$, and $V_{x,3}$.

3. Clearly, $\{b_0\}$ is a basis of $W_{x,1}$ and $\{b_{00}\}$ is a basis of $V_{x,1}$. Also, any element of $W_{x,2}$ and $V_{x,2}$ can be written as a linear combination of "vectors" in $\{b_r;\ r = 1,2,3\}$ and $\{b_{0r};\ r = 1,2,3\}$, respectively. Finally, the set $\{b_{rs};\ r,s = 1,2,3\}$ is a system of generators for $V_{x,3}$. To make $\mathcal{W}_x$ into an inner product space (Greub, 1975), we determine the scalar product $\langle \cdot , \cdot \rangle_x$ in $\mathcal{W}_x$ such that the following conditions hold:

$$\langle b_0, b_0 \rangle_x := -g_x^{\mathrm{RW}}(b_0, b_0) = 1 \tag{3.35a}$$

$$\langle b_r, b_s \rangle_x := g_x^{\mathrm{RW}}(b_r, b_s) = R^2(t)\delta_{rs} \tag{3.35b}$$

$$\langle b_{00}, b_{00} \rangle_x := [-g_x^{\mathrm{RW}}(b_0, b_0)]^2 = 1 \tag{3.36a}$$

$$\langle b_{0r}, b_{0s} \rangle_x := -\tfrac{1}{2} g_x^{\mathrm{RW}}(b_0, b_0) g_x^{\mathrm{RW}}(b_r, b_s) = \tfrac{1}{2} R^2(t)\delta_{rs} \tag{3.36b}$$

$$\begin{aligned} \langle b_{ij}, b_{rs} \rangle_x &:= \tfrac{1}{2}[g_x^{\mathrm{RW}}(b_i, b_r) g_x^{\mathrm{RW}}(b_j, b_s) + g_x^{\mathrm{RW}}(b_i, b_s) g_x^{\mathrm{RW}}(b_j, b_r)] \\ &= \tfrac{1}{2} R^4(t)(\delta_{ir}\delta_{js} + \delta_{is}\delta_{jr}) \end{aligned} \tag{3.36c}$$

Now, it is important to observe that equations (3.35) and (3.36) are not independent, because the scalar products in $W_{x,1}$ and $W_{x,2}$ induce the scalar products in $V_{x,1}$, $V_{x,2}$, and $V_{x,3}$. [By way of digression, an analogous problem which has been solved was mentioned by Emch (1972) in his discussion of the properties of the scalar product for Fock space.]

4. We have then, up to here, constructed the scalar products for $W_{x,1}$, $W_{x,2}$, $V_{x,1}$, $V_{x,2}$ and $V_{x,3}$. As regards R_1 and R_2 ($\mathrm{R}_1 = \mathrm{R}_2 = \mathrm{R}$), the set of real numbers R is a vector space together with a binary law of composition $\langle \cdot , \cdot \rangle_x$: $\mathrm{R} \times \mathrm{R} \Rightarrow \mathrm{R}$ defined by

$$\langle y, z \rangle_x := y \cdot z, \qquad y, z \in \mathrm{R} \tag{3.37}$$

In this way R_1 and R_2 become inner product spaces as well.

5. Starting from these definitions and using the decomposition

$$\mathcal{W}_x = V_{x,1} \oplus V_{x,2} \oplus V_{x,3} \oplus W_{x,1} \oplus W_{x,2} \oplus \mathrm{R}_1 \oplus \mathrm{R}_2 \tag{3.38}$$

we can easily introduce a natural scalar product in $\mathcal{W}_x$. The general idea is in fact quite simple (Greub, 1975). Consider first the inner product spaces $V_{x,1}$ and $V_{x,2}$. Prove that a scalar product is given in the direct sum $V_{x,1} \oplus V_{x,2}$ by

$$\langle y_1 \oplus y_2, z_1 \oplus z_2 \rangle_x := \langle y_1, z_1 \rangle_x + \langle y_2, z_2 \rangle_x \tag{3.39}$$

where y_1, $z_1 \in V_{x,1}$ and y_2, $z_2 \in V_{x,2}$. This construction can now be repeated for $(\mathrm{V}_{x,1} \oplus V_{x,2}) \oplus V_{x,3}$ with the inner product spaces $V_{x,1} \oplus V_{x,2}$ and $V_{x,3}$ playing the role previously played by $V_{x,1}$ and $V_{x,2}$. In the last step we obtain the bilinear function $\langle y, z \rangle_x$, which is a scalar product in $\mathcal{W}_x$.

These remarks end our discussion concerning the definition of a scalar product in $\mathcal{W}_x$. As we shall see in Sections 4 and 5, the existence of $\langle y, z \rangle_x$ and similar objects ensures the possibility of a systematic derivation of the countably infinite set of gauge-invariant variables.

4. SOME ELEMENTARY GAUGE-INVARIANT QUANTITIES

4.1. Generalities

With the scalar product of Section 3.4, we may ask now if there is some natural procedure for introducing the notion of a gauge-invariant quantity. It turns out that we can define the gauge-invariant quantities as follows. Let $\mathcal{H}_x$ be the set of $h \in \mathcal{W}_x$ such that the scalar product $\langle h, \mathcal{L}_v C_0 \rangle_x$ vanishes for all differentiable vector fields v on X:

$$\mathcal{H}_x := \{h \in \mathcal{W}_x : \bigwedge_v \langle h, \mathcal{L}_v C_0 \rangle_x = 0\} \tag{4.1}$$

Clearly, $\mathcal{H}_x$ is a subspace of $\mathcal{W}_x$. Given the bilinear function $\langle \cdot, \cdot \rangle_x$, we denote by $\mathcal{F}_x$ the set of all "vectors" which are orthogonal to $\mathcal{H}_x$. $\mathcal{F}_x$ is again a subspace of $\mathcal{W}_x$ and the intersection $\mathcal{H}_x \cap \mathcal{F}_x$ consists of the zero-vector only. As a corollary, we have the decomposition

$$\mathcal{W}_x = \mathcal{H}_x \oplus \mathcal{F}_x \tag{4.2}$$

Put

$$\mathcal{H} := \bigcup_{x \in X} \mathcal{H}_x \tag{4.3}$$

A glance at equation (3.14) shows that $\mathcal{H}$ is a vector subbundle of $\mathcal{W}$. Let $h: X \Rightarrow \mathcal{H}$ be a cross section of $\mathcal{H}$ (this cross section is not necessarily continuous), and suppose that [P] is a gauge-invariant perturbation associated with P. If $\mathrm{P}' \in [\mathrm{P}]$, one then obtains

$$\langle h, \mathrm{P} \rangle_x = \langle h, \mathrm{P}' \rangle_x \tag{4.4}$$

An equivalent statement is that

$$J_h(x) := \langle h, \mathrm{P} \rangle_x \tag{4.5}$$

represents a *gauge-invariant quantity* induced by h.

What are the simplest gauge-invariant quantities in an almost-Robertson–Walker universe model? Since dim $\mathcal{H}_x = 2$, we can easily determine them by specifying two elementary cross sections h_1 and h_2 of $\mathcal{H}$ such that $\{h_1(x), h_2(x)\}$ forms a basis of $\mathcal{H}_x$ for each $x \in X$, and then describing the action of $\langle h_1, \cdot \rangle_x$ and $\langle h_2, \cdot \rangle_x$ on $\mathrm{P}(x)$:

$$\chi(x) := J_1(x) := \langle h_1, \mathrm{P} \rangle_x \tag{4.6a}$$

$$\Gamma(x) := J_2(x) := \langle h_2, \mathrm{P} \rangle_x \tag{4.6b}$$

However, to carry this out, we must first construct a moving frame for the vector bundle $\mathcal{F} := \bigcup_{x \in X} \mathcal{F}_x$ over X.

4.2. Specification of χ and Γ

Consider equation (3.30) for $(\mathcal{L}_v C_0)(x)$. Since the values of v^α and $v^\alpha_{,\beta}$ can be chosen arbitrarily at a fixed point $x \in X$, each "vector" $(\mathcal{L}_v C_0)(x)$ is to be a linear combination of f_0, f_{0r}, f_{rs}, f_r and f $(r,s = 1,2,3)$; by definition, we have

$$f_0 := -2b_{00} \oplus b_0 \in V_{x,1} \oplus W_{x,1} \subset \mathcal{W}_x \tag{4.7a}$$

$$f_{0r} := R^{-1} b_{0r} \in V_{x,2} \subset \mathcal{W}_x \tag{4.7b}$$

$$f_{rs} := R^{-2} b_{rs} \in V_{x,3} \subset \mathcal{W}_x \tag{4.7c}$$

$$f_r := R^{-1} b_r \in W_{x,2} \subset \mathcal{W}_x \tag{4.7d}$$

$$f := 1 \oplus (-(3HT_0)^{-1}\dot{T}_0) \in \mathbf{R}_1 \oplus \mathbf{R}_2 \subset \mathcal{W}_x \tag{4.7e}$$

In fact, it follows from the above definitions that the vectors $f_0, f_{0r}, f_{rs}, f_r, f$ are *mutually orthogonal* and hence they form a basis of $\mathcal{F}_x$. The dimension of $\mathcal{F}_x$ is therefore 14. Since dim $\mathcal{W}_x = 16$, we then find that

$$\dim \mathcal{H}_x = \dim \mathcal{W}_x - \dim \mathcal{F}_x = 16 - 14 = 2 \tag{4.8}$$

Passing over to the orthogonal complement of $\mathcal{F}_x$, we obtain a basis $\{h_1(x), h_2(x)\}$ of $\mathcal{H}_x$ by the Schmidt-orthogonalization process (Szegö, 1939)

$$h_1 := b_{00} \oplus 2b_0 \in V_{x,1} \oplus W_{x,1} \subset \mathcal{W}_x \tag{4.9a}$$

$$h_2 := (3HT_0)^{-1}\dot{T}_0 \oplus 1 \in \mathbf{R}_1 \oplus \mathbf{R}_2 \subset \mathcal{W}_x \tag{4.9b}$$

We can now use equations (3.20), (3.21), and (4.6) to compute the gauge-invariant quantities $\chi(x)$ and $\Gamma(x)$ in the usual fashion, showing that

$$\chi(x) = Q^{00}(x) + 2U^0(x) \tag{4.10a}$$

and

$$\Gamma(x) = K(x) + (3HT_0)^{-1}\dot{T}_0 M(x) \tag{4.10b}$$

An equation for $\chi(x)$ can also be obtained from the condition $u^\alpha(\epsilon,x)u_\alpha(\epsilon,x) = -1$, with the result

$$\chi = -\left(\frac{\partial(u^\alpha u_\alpha)}{\partial \epsilon}\right)_{|\epsilon=0} = 0 \tag{4.11}$$

Thus the gauge-invariant quantity $\chi(x)$ will not be physically significant to us in considering linearization about the Robertson–Walker universe models. This conclusion, however, does not mean that the identity $\chi = 0$ is not mathematically important; it will be used in Section 6. As to the physical interpretation of Γ, in a companion paper (Banach and Piekarski, 1996) we prove that *this quantity is proportional to the entropy perturbation* δs.

Because of equation (4.2), one is of course free to write the infinitesimal perturbation $\mathrm{P}(x)$ at $x \in X$ as a sum of two distinctive parts: the first part represents the orthogonal projection of $\mathrm{P}(x)$ onto $\mathcal{F}_x$, whereas the second part lies in $\mathcal{H}_x$ and is given by

$$\mathrm{P}_H(x) = \frac{1}{5}\chi(x)h_1 + \left[1 + \left(\frac{1}{3HT_0}\dot{T}_0\right)^2\right]^{-1}\Gamma(x)h_2$$

$$= \left[1 + \left(\frac{1}{3HT_0}\dot{T}_0\right)^2\right]^{-1}\Gamma(x)h_2 \tag{4.12}$$

Since there is a natural scalar product in $\mathcal{W}_x$, there is also a natural norm by which "smallness" of $\mathrm{P}(x)$ can be measured. An adequate definition of "smallness" in this context is that

$$\|\mathrm{P}(x)\| := (\langle \mathrm{P}, \mathrm{P}\rangle_x)^{1/2} \tag{4.13a}$$

or

$$\|\mathrm{P}_H(x)\| := (\langle \mathrm{P}_H, \mathrm{P}_H\rangle_x)^{1/2} \tag{4.13b}$$

be much smaller than 1 in some region of space-time.

Finally, one important point should be noticed. The norm $\|\mathrm{P}_H(x)\|$ is completely independent of the gauge chosen. However, the norm of $\mathrm{P}(x)$ is not gauge invariant: it can be assigned any value we like at any event by an appropriate choice of the vector field v in $\mathrm{P}(x) + (\mathcal{L}_v C_0)(x)$.

4.3. Difficulties with the Simple Theory

As is clear from the discussion so far in this paper, the gauge-invariant variables χ and Γ are not sufficient to represent the "point" [P] of the quotient space $\mathcal{P}/\mathcal{P}_0$ explicitly. One could have circumvented these sorts of difficulties altogether if one had chosen, at the outset, to consider carefully the degrees of freedom associated with the first-, second-, and higher order covariant derivatives of $g(\epsilon,x)$, $u(\epsilon,x)$, $n(\epsilon,x)$, and $T(\epsilon,x)$ with respect to g_x^{RW}. In that case, one would have been led to introduce as the basic object of one's theory the generalization of the notion of an infinitesimal perturbation at a point $x \in X$, which contains more information about the behavior of the fields Q, U, M, and K in a neighborhood of x. By means of such a generalization, one could then proceed with complete rigor in deriving many further gauge-invariant variables for an almost-Robertson–Walker universe model. This is the objective of Section 5.

5. PROLONGATION OF THE ORIGINAL STRUCTURES

5.1. Differentiation of the Number Density

In physical cosmology (Pebbles, 1993), the most interesting gauge-invariant quantities are those involving the temporal and spatial gradients of the number density $n(\epsilon,x)$. We can represent these gradients by

$$n'_\alpha(\epsilon, x) := \frac{1}{\dot{n}_0(t)} \frac{\partial n(\epsilon, x)}{\partial x^\alpha} \tag{5.1}$$

Because of equations (2.8c), we find that

$$M'_\alpha(x) := \left(\frac{\partial n'_\alpha(\epsilon, x)}{\partial \epsilon}\right)_{|\epsilon=0} = \frac{1}{\dot{n}_0(t)} \frac{\partial(n_0(t)M(x))}{\partial x^\alpha} \tag{5.2}$$

Let $\{b^\alpha\}$ be a basis of one-forms dual to a coordinate basis $\{b_\alpha\}$. Then $n'_\alpha(\epsilon,x)$ are the components of

$$n'(\epsilon, x) := n'_\alpha(\epsilon, x)b^\alpha \in W^*_x \tag{5.3}$$

with respect to $\{b^\alpha\}$; here W^*_x denotes the vector space dual to W_x. Also, it will be convenient to set

$$M'(x) := M'_\alpha(x)b^\alpha \in W^*_x \tag{5.4}$$

Clearly, the decomposition $W_x = W_{x,1} \oplus W_{x,2}$ implies that W^*_x is a direct sum of two dual subspaces $W^*_{x,1}$ and $W^*_{x,2}$:

$$W^*_x := W^*_{x,1} \oplus W^*_{x,2} \tag{5.5}$$

Using equation (3.6), we define the scalar product $\langle y,z\rangle_x$ in W^*_x by

$$\langle b^0, b^0\rangle_x := -g_{(0)x}(b^0, b^0) = 1 \tag{5.6a}$$

and

$$\langle b^r, b^s\rangle_x := g_{0(x)}(b^r, b^s) = R^{-2}(t)\delta^{rs} \tag{5.6b}$$

If y and z are general linear combinations of b^0 and b^r, the value of $\langle y,z\rangle_x$ can then be calculated explicitly and expressed in terms of $\langle b^0,b^0\rangle_x$ and $\langle b^r,b^s\rangle_x$ [see Problem 4 in Greub (1975, p. 191)]. In this way W^*_x becomes an inner product space. Finally, we shall simply write

$$\mathcal{W}_x := V_x \oplus W_x \oplus \mathbb{R}_1 \oplus \mathbb{R}_2 \oplus W^*_x \tag{5.7}$$

and it is to be understood that $\mathcal{W}_x$ is an inner product space as well.

We continue to denote the infinitesimal perturbation at a point $x \in X$ by $\mathrm{P}(x)$, but it is now regarded as a member of $V_x \oplus W_x \oplus \mathbf{R}_1 \oplus \mathbf{R}_2 \oplus W_x^*$, and defined by the following condition:

$$\mathrm{P}(x) := \left(\frac{\partial C(\epsilon, x)}{\partial \epsilon}\right)_{|\epsilon=0} \tag{5.8}$$

in which $C(\epsilon,x)$ stands for

$$C(\epsilon,x) := g(\epsilon,x) \oplus u(\epsilon,x) \oplus \bar{n}(\epsilon,x) \oplus \overline{T}(\epsilon,x) \oplus n'(\epsilon,x) \tag{5.9}$$

Hence

$$\mathrm{P}(x) = Q(x) \oplus U(x) \oplus M(x) \oplus K(x) \oplus M'(x) \tag{5.10}$$

The counterpart of equation (3.27) is then

$$\mathcal{L}_v C_0 := \mathcal{L}_v g_{(0)} \oplus \mathcal{L}_v u_{(0)} \oplus \overline{\mathcal{L}_v n_0} \oplus \overline{\mathcal{L}_v T_0} \oplus (\mathcal{L}_v n_0)' \tag{5.11}$$

where

$$(\mathcal{L}_v n_0)' := (\dot{n}_0)^{-1}(\mathcal{L}_v n_0)_{,\alpha} b^\alpha = (\dot{n}_0)^{-1}(v^0 \dot{n}_0)_{,\alpha} b^\alpha \tag{5.12}$$

As in equation (4.2), we can obtain a decomposition

$$\mathcal{W}_x = \mathcal{H}_x \oplus \mathcal{F}_x \tag{5.13}$$

such that all the definitions and all the observations of Section 4.1, with the exception of the interpretation of $\mathcal{L}_v C_0$, remain valid when we replace equation (3.27) by equation (5.11) in the statements and proofs.

It is readily confirmed by substituting for $\mathcal{L}_v g_{(0)}$, $\mathcal{L}_v u_{(0)}$, $\overline{\mathcal{L}_v n_0}$, $\overline{\mathcal{L}_v T_0}$, and $(\mathcal{L}_v n_0)'$ from equations (3.24), (3.28), and (5.12) into equation (5.11) that $(\mathcal{L}_v C_0)(x)$ takes the form

$$\begin{aligned}(\mathcal{L}_v C_0)(x) = {} & -3v^0 H\left[\left(\frac{2}{3} R^{-2}\delta^{rs} b_{rs}\right) \oplus 1 \oplus (-(3HT_0)^{-1}\dot{T}_0)\right. \\ & \left.\oplus (-(3H\dot{n}_0)^{-1}(n_0)^{\cdot\cdot} b^0)\right] - v^0_{,0}[-2b_{00} \oplus b_0 \oplus (-b^0)] \\ & - (\delta^{rp} v^s_{,p} + \delta^{sp} v^r_{,p})(R^{-2} b_{rs}) \\ & + \frac{1}{2}\left(\frac{1}{R}\delta^{rp} v^0_{,p} - R v^r_{,0}\right)\left[\left(-\frac{4}{R} b_{0r}\right) \oplus \left(\frac{1}{R} b_r\right) \oplus (R\delta_{rs} b^s)\right] \\ & + \frac{1}{2}\left(\frac{1}{R}\delta^{rp} v^0{}_{,p} + R v^r_{,0}\right)\left[\left(-\frac{1}{R} b_r\right) \oplus (R\delta_{rs} b^s)\right]\end{aligned} \tag{5.14}$$

with summation over r, s, p = 1,2,3 assumed, so the required basis $\{h_1, h_2, h_0, h_{0r}\}$ of $\mathcal{H}_x$ may be written as

$$h_1 := b_{00} \oplus 2b_0 \in V_{x,1} \oplus W_{x,1} \subset \mathcal{W}_x \tag{5.15a}$$

$$h_2 := (3HT_0)^{-1}\dot{T}_0 \oplus 1 \in \mathbf{R}_1 \oplus \mathbf{R}_2 \subset \mathcal{W}_x \tag{5.15b}$$

$$h_0 := -\frac{1}{2}\, b_{00} \oplus (3H\dot{n}_0)^{-1}(n_0)^{\cdot\cdot} \oplus b^0 \in V_{x,1} \oplus \mathbf{R}_1 \oplus W^*_{x,1} \subset \mathcal{W}_x \tag{5.15c}$$

$$h_{0r} := \frac{1}{R}\, b_{0r} \oplus \frac{1}{R}\, b_r \oplus Rb^r \in V_{x,2} \oplus W_{x,2} \oplus W^*_{x,2} \subset \mathcal{W}_x \tag{5.15d}$$

where r = 1, 2, 3. This basis is not orthogonal. However, starting out with $\{h_1, h_2, h_0, h_{0r}\}$, a new basis can be constructed whose vectors are mutually orthogonal. Inspection shows that dim $\mathcal{H}_x = 6$ and dim $\mathcal{F}_x = 14$; thus dim $\mathcal{W}_x = 20$.

Once the basis of $\mathcal{H}_x$ has been put into the explicit form, the construction of the gauge-invariant quantities

$$\chi(x) := J_1(x) := \langle h_1, \mathrm{P}\rangle_x \tag{5.16a}$$

$$\Gamma(x) := J_2(x) := \langle h_2, \mathrm{P}\rangle_x \tag{5.16b}$$

$$\Omega(x) := J_0(x) := \langle h_0, \mathrm{P}\rangle_x \tag{5.16c}$$

$$\Omega^r(x) := -3RHJ_{0r}(x) := -3RH\langle h_{0r}, \mathrm{P}\rangle_x \tag{5.16d}$$

is extremely simple, for the expressions in equations (5.16) become

$$\chi(x) = Q^{00}(x) + 2U^0(x) \tag{5.17a}$$

$$\Gamma(x) = K(x) + (3HT_0)^{-1}\dot{T}_0M(x) \tag{5.17b}$$

$$\Omega(x) = -\frac{1}{2}\, Q^{00}(x) + \frac{1}{3}\, H^{-2}[\dot{H}M(x) - H\dot{M}(x)] \tag{5.17c}$$

$$\Omega^r(x) = -3R^2HQ^{0r}(x) - 3R^2HU^r(x) + \delta^{rs}\,\frac{\partial M}{\partial x^s} \tag{5.17d}$$

Here it is perhaps important to recall that $Q^{\alpha\beta}(x)$ and $U^\alpha(x)$ are related to $Q(x)$ and $U(x)$ by equations (3.21a) and (3.21b) [see also equation (2.8a) for the definition of $Q^{\alpha\beta}(x)$ and $U^\alpha(x)$]. Clearly, in order to obtain equations (5.17) from equations (5.16), we have used

$$\dot{n}_0 = -3n_0H \tag{5.18a}$$

and

$$(n_0)^{\cdot\cdot} = -3n_0\dot{H} + 9n_0H^2 \tag{5.18b}$$

We stress that the objects χ, Γ, Ω, and Ω^r are gauge-invariant quantities. The physical importance of equations (5.17) follows from the fact that they are all necessary to give a tractable and explicit description of the elements P of a quotient space $\mathcal{P}/\mathcal{P}_0$; this is the problem of most fundamental interest. In a companion paper (Banach and Piekarski, 1996) we demonstrate that, to first order in the deviations from the background solution, *the quantity* $n_0R^{-2}\Omega^r$ *represents the gauge-invariant spatial gradient of the number density* n. However, the validity or utility of our definitions does not depend on this interpretation. As an illustration, the gauge-invariant object Ω has a more remote physical significance; in a synchronous gauge ($Q^{00} = Q^{0r} = 0$) it gives us information about the time variation of the density contrast M. Nevertheless, this object is absolutely necessary to define a "coordinate system" for the unique description of $\mathcal{P}/\mathcal{P}_0$.

It is always possible to decompose our basic variables harmonically, thus separating out the time and space variations; this is done in Banach and Piekarski (1996, Section 6), where comparisons with the interesting methods of Bardeen (1980) and Ellis and Bruni (1989) are also made.

5.2. A Complete Set of Basic Gauge-Invariant Quantities

In Section 5.1, we obtained two independent gauge-invariant quantities Ω and Ω^r that code the information we need to discuss density inhomogeneities in an almost-Robertson–Walker universe model. The starting point for the analysis was a "prolonged" vector space $V_x \oplus W_x \oplus \mathbf{R}_1 \oplus \mathbf{R}_2 \oplus W_x^*$ together with an appropriate scalar product $\langle y,z \rangle_x$. Since Ω and Ω^r are linear in first derivatives of the density contrast M, one might ask whether modifying a definition of the prolonged vector space might not give an extended set of gauge-invariant variables. Such is indeed the case. However, for lack of space we will not present here all the details concerning this modification, but suffice it to say that if a suitable hierarchy of prolonged vector spaces is known, a countably infinite set of gauge-invariant quantities can be systematically constructed. In fact, given this hierarchy, it will be possible to find many further gauge-invariant quantities by using the standard procedure of orthogonalization (the Schmidt orthogonalization). Among the problems that can be studied with this sort of approach, an explicit description of the elements of the quotient space $\mathcal{P}/\mathcal{P}_0$ presents a most interesting challenge (see Section 6). However, before proceeding to examine this issue, we first introduce a complete set of basic gauge-invariant variables. (The details of the construction, which are very much analogous to those of Section 5.1, are available on request.)

We start our discussion as follows. In an almost-Robertson–Walker universe model, one can choose coordinates so that the unperturbed metric tensor g_x^{RW} has the form (Misner *et al.*, 1973)

$$g_x^{\mathrm{RW}} = -b^{00} + [R(t)/\omega(x)]^2\delta_{rs}b^{rs} \qquad (5.19a)$$

where

$$\omega(x) := 1 + (k/4)[(x^1)^2 + (x^2)^2 + (x^3)^2] \qquad (5.19b)$$

and where k is the (constant) spatial curvature. By an appropriate choice of units, the value of k can be made to be $+1$, -1, or 0. The corresponding solutions for the expansion factor $R(t)$ represent, respectively, spaces of positive or negative curvature or flat space. The tensor γ_x given by

$$\gamma_x := [1/\omega(x)]^2\delta_{rs}b^{rs} = \gamma_{rs}b^{rs} \qquad (5.20)$$

is the metric tensor of a three-space of uniform spatial curvature k. *Let a slash denote the covariant derivative of a three-tensor with respect to* γ_x. The components of γ_x are $\gamma_{rs} = (1/\omega)^2\delta_{rs}$. Knowing γ_{rs}, we define the contravariant tensor γ^{rs} by $\gamma^{rs} = \omega^2\delta^{rs}$.

With all these definitions in mind, a complete set of basic gauge-invariant variables is given by

$$\chi := Q^{00} + 2U^0 \qquad (5.21a)$$

$$\Gamma := K + (3HT_0)^{-1}\dot{T}_0 M \qquad (5.21b)$$

$$\Omega := -\tfrac{1}{2}Q^{00} + \tfrac{1}{3}H^{-2}(\dot{H}M - H\dot{M}) \qquad (5.21c)$$

$$\Omega^r = -3R^2HQ^{0r} - 3R^2HU^r + \gamma^{rs}M_{|s} \qquad (5.21d)$$

$$\Delta := -\frac{3}{2}Q^{00} - R^2\gamma_{rs}\left(Q^{rs} + \frac{1}{2H}\dot{Q}^{rs}\right) + \frac{1}{H}U^r{}_{|r} + H^{-2}\dot{H}M \qquad (5.21e)$$

$$\Delta^{rs} := R^2\left(Q^{rs} + \frac{1}{2H}\dot{Q}^{rs}\right) - \frac{1}{2H}(\gamma^{rp}U^s{}_{|p} + \gamma^{sp}U^r{}_{|p})$$
$$- \frac{1}{3}R^2\gamma_{pq}\left(Q^{pq} + \frac{1}{2H}\dot{Q}^{pq}\right)\gamma^{rs} + \frac{1}{3H}U^p{}_{|p}\gamma^{rs} \qquad (5.21f)$$

$$S^{ijrs} := kZ^{p[i}(\gamma^{j]r}\delta^s_p - \gamma^{j]s}\delta^r_p) + \gamma^{sq}\gamma^{p[i}Z^{j]r}{}_{|pq} - \gamma^{rq}\gamma^{p[i}Z^{j]s}{}_{|pq} \qquad (5.21g)$$

where δ^r_p is the Kronecker delta, $M_{|s} = M_{,s}$, and

$$Z^{rs} := -R^2Q^{rs} + \frac{2}{3}M\gamma^{rs} \qquad (5.22)$$

Here the antisymmetric part of $\mathcal{M}^{ij\ldots s}$ is distinguished by square brackets.

Inspection shows that χ vanishes identically and that $\gamma_{rs}\Delta^{rs} = 0$. A complete set of symmetry conditions for S^{ijrs} is $S^{ijrs} = S^{[ij][rs]}$ and $S^{i[jrs]} = 0$; thus there are six linearly independent, not identically vanishing components in $\{S^{ijrs} | i, j, r, s = 1, 2, 3\}$.

The gauge-invariant objects Δ and S^{ijrs} have no direct physical meaning, except for the fact that they are constructed from the metric and matter perturbations. As regards the traceless three-tensor Δ^{rs}, it is possible to show that, to first order in the deviations from the background solution, *this gauge-invariant quantity has a direct physical interpretation in terms of the shear of the matter velocity field* (Banach and Piekarski, 1996). The basic role of equations (5.21) is in essence that of a guarantor for the existence of a unique and transparent description of $\mathcal{P}/\mathcal{P}_0$ (see Section 6).

In the present approach, we proceeded in a manner similar to the process for diagonalizing a quadratic function, obtaining coordinates (x^α) for which the background metric and the gauge-invariant tensor fields have as simple a form as possible. Since symmetry is broken by the existence of a background solution for the fluid four-velocity u, our formulation of perturbation theory in terms of a privileged system of coordinates should be considered a substantial labor-saving device. Of course, the construction of (x^α) is not unique, because we can easily introduce new coordinate systems on X for which the decomposition (5.19a) holds, once one choice of (x^α) has been made.

6. EXPLICIT DESCRIPTION OF THE QUOTIENT SPACE

Given a complete set of basic gauge-invariant variables, we can now formulate our main theorem. From this theorem it follows that two infinitesimal perturbations $\delta\mathcal{G}_0$ and $\delta\mathcal{G}_0'$ are equivalent if (and only if) they determine *one common set of basic gauge-invariant variables*, i.e., if (and only if) D = D′, where D denotes a basic set [see equation (6.10) for the definition of D].

Theorem. Let $Q^{\alpha\beta}$, U^α, M, and K be functions of class C^r (r sufficiently large) on a differentiable manifold X, and suppose that these functions obey the following conditions:

$$\chi = 0, \qquad \Gamma = 0, \qquad \Omega = 0, \qquad \Delta = 0 \tag{6.1a}$$

$$\Omega^r = 0, \qquad \Delta^{rs} = 0, \qquad S^{ijrs} = 0 \tag{6.1b}$$

Then there exists a vector field v^α on X such that

$$Q^{00} = 2\dot{v}^0 \tag{6.2a}$$

$$Q^{0r} = -R^{-2}\gamma^{rp}v^0{}_{|p} + \dot{v}^r, \qquad v^0{}_{|p} = v^0_{,p} \tag{6.2b}$$

$$Q^{rs} = -2R^{-2}Hv^0\gamma^{rs} - R^{-2}(\gamma^{rp}v^s{}_{|p} + \gamma^{sp}v^r_{|p}) \tag{6.2c}$$

$$U^\alpha = -\dot{v}^\alpha \tag{6.2d}$$

$$M = \frac{1}{n_0} v^0 \dot{n}_0 = -3Hv^0 \tag{6.2e}$$

$$K = \frac{1}{T} v^0 \dot{T}_0 \tag{6.2f}$$

Remark. Interpreting this theorem, equations (6.1) and (6.2) imply that the infinitesimal perturbation $\delta\mathscr{G}_0 := (Q^{\alpha\beta}, U^\alpha, n_0 M, T_0 K)$ can be identified with the Lie derivative

$$\mathscr{L}_v\mathscr{G}_0 := ((\mathscr{L}_v g_{(0)})^{\alpha\beta}, (\mathscr{L}_v u_{(0)})^\alpha, \mathscr{L}_v n_0, \mathscr{L}_v T_0)$$

of the background solution $\mathscr{G}_0 := ((g_{(0)})^{\alpha\beta}, (u_{(0)})^\alpha, n_0, T_0)$ with respect to the aforementioned vector field v. Clearly, such a perturbation is a "gauge mode" solution of equation (2.10).

Sketch of the proof. Write

$$v^0 := (\dot{n}_0)^{-1} n_0 M = -\frac{1}{3H} M \tag{6.3}$$

and define the functions v^1, v^2, and v^3 by saying that they satisfy the differential equations of the form

$$\dot{v}^r = -U^r, \qquad r = 1, 2, 3 \tag{6.4}$$

This trivially proves equation (6.2e) and equation (6.2d) for $\alpha = 1, 2, 3$. Because of $\Gamma = 0$ and $\Omega = 0$, K equals $T_0^{-1} v^0 \dot{T}_0$, Q^{00} equals $2\dot{v}^0$, and equations (6.2f) and (6.2a) hold. Then we may conclude from $\chi = 0$ that equation (6.2d) is valid for $\alpha = 0$ as well. After substituting equations (6.2d) and (6.2e) into the left-hand side of $\Omega^r = 0$, we immediately arrive at equations (6.2b). Given the definition (6.3) of v^0, we now set [see equations (5.22) and (6.2e)]

$$Z^{rs} := -R^2 Q^{rs} + \frac{2}{3} M \gamma^{rs} = -R^2 Q^{rs} - 2Hv^0 \gamma^{rs} \tag{6.5}$$

Hence the conditions $\Delta = 0$ and $\Delta^{rs} = 0$ take the form

$$\dot{Z}^{rs} = \gamma^{rp} \dot{v}^s{}_{|p} + \gamma^{sp} \dot{v}^r{}_{|p} \tag{6.6}$$

Because of $\dot{\omega} = 0$, this is equivalent to

$$Z^{rs} = \gamma^{rp} v^s{}_{|p} + \gamma^{sp} v^r{}_{|p} + \upsilon^{rs} \tag{6.7}$$

where $\{\upsilon^{rs}\}$ is the second-rank, symmetric three-tensor whose components

υ^{rs} are independent of time. Further, as an explicit application of the condition $S^{ijrs} = 0$, we derive the differential equations for $\upsilon^{rs}(x^1, x^2, x^3)$:

$$k\upsilon^{p[i}(\gamma^{j]r}\delta^s_p - \gamma^{j]s}\delta^r_p) + \gamma^{sq}\gamma^{p[i}\upsilon^{j]r}{}_{|pq} - \gamma^{rq}\gamma^{p[i}\upsilon^{j]s}{}_{|pq} = 0 \tag{6.8}$$

However, equation (6.8) is a *necessary* and *sufficient* condition [Truesdell and Toupin (1960), equation (84.12), p. 352] that, given a symmetric tensor $\upsilon^{rs}(x^1, x^2, x^3)$, there exists a vector field $\upsilon^r(x^1, x^2, x^3)$ such that[3]

$$\upsilon^{rs} = \gamma^{rp}\upsilon^s{}_{|p} + \gamma^{sp}\upsilon^r{}_{|p} \tag{6.9}$$

Of course, since $\dot{v}^r = \dot{v}^r + \dot{\upsilon}^r = -U^r$, we may always choose to replace v^r by $v^r + \upsilon^r$. Hence if we suppose the vector field v^r so adjusted that $\upsilon^r = 0$, by combining equations (6.5), (6.7), and (6.9) we obtain equation (6.2c). This completes the proof of our theorem. ∎

To go into further detail regarding a gauge-invariant formulation of perturbation theory would take more space than is available here and would yield the conclusions and observations which we present in a companion paper (Banach and Piekarski, 1996). The outline already given satisfies the purpose of this paper: to show that a complete set of basic gauge-invariant variables, applied to an almost-Robertson–Walker universe model ($k = -1$, 0, $+1$), suffices to obtain an explicit description of $[\mathrm{P}] \in \mathcal{P}/\mathcal{P}_0$. In fact, there is something very special about the way the basic set

$$\mathrm{D} := \{\chi, \Gamma, \Omega, \Omega^r, \Delta, \Delta^{rs}, S^{ijrs}\} \tag{6.10}$$

is related to the gauge-invariant perturbation [P], for we can *uniquely* recover [P] from D and conversely.

To see this, it is easiest to start with two infinitesimal perturbations P and P′ (P ≠ P′) such that D = D′, where D and D′ are the corresponding sets of basic variables. (Here we recall that these infinitesimal perturbations are solutions of the gauge-dependent propagation equations). However, according to our theorem, in linear perturbation theory the relation D = D′ is equivalent to saying that $\mathrm{P}' = \mathrm{P} + \mathcal{L}_v C_0$ for some v ($v \neq 0$). Hence P and P′ differ by the action of an "infinitesimal diffeomorphism" on the background solution and any fixed set D can be used to extract [P] from D in a unique way. This in turn means that D "is" a gauge-invariant perturbation:

$$[\mathrm{P}] \Leftrightarrow \mathrm{D} \tag{6.11}$$

In other words, the correspondence (6.11) defines a bijection A from $\mathcal{P}/\mathcal{P}_0$ onto the "vector space" $\mathcal{D}$ such that if $[\mathrm{P}] \in \mathcal{P}/\mathcal{P}_0$, then $A([\mathrm{P}]) =$

[3]The notation of Truesdell and Toupin (1960) differs from that of ours as follows: they denote by $\mathcal{M}^{ij\ldots r}{}_{,s}$ the covariant derivative of $\mathcal{M}^{ij\ldots r}$ with respect to χ_{rs}.

D is a set of basic gauge-invariant variables associated with [P]. We call $\mathfrak{D}$ the image space of $\mathcal{P}/\mathcal{P}_0$ under A or simply the *image space* of $\mathcal{P}/\mathcal{P}_0$. Clearly, from these observations we infer that D := $\{\chi, \Gamma, \Omega, \Omega^r, \Delta, \Delta^{rs}, S^{ijrs}\}$ belongs to $\mathfrak{D}$ if (and only if) there is an infinitesimal perturbation P or $\delta\mathcal{G}_0$ satisfying equations (2.10) and (5.21) with Z^{rs} given by (5.22); *since A is a bijection, such perturbations always exist.* This definition of $\mathfrak{D}$ will be significant to us (Banach and Piekarski, 1996) in interpreting some of the solutions of gauge-invariant equations for gauge-invariant variables. Finally, we would like to mention that D contains a maximum number of independent variables, and any other gauge-invariant variable can be expressed in terms of D (see the Appendix). The total number of linearly independent, not identically vanishing components in D is 17.

7. DISCUSSION AND CONCLUSION

We have presented a totally gauge-invariant framework for studying perturbation theory away from homogeneous and isotropic cosmological models. The principal new results of this paper are contained in Sections 5.2 and 6. Since physical perturbations can always be regarded as being the elements of a certain quotient space, the essential ingredient in the discussion was the definition and construction of a complete set of basic gauge-invariant variables in terms of which to describe the elements of this quotient space explicitly. An additional welcome feature is that, as shown in the Appendix, any other gauge-invariant quantity can be determined directly from these basic variables through purely algebraic and differential operations. The calculations were presented in some detail for the special case of a perturbed $k = 0$ Robertson–Walker space-time. The generalization to a three-space of uniform spatial curvature k ($k = -1, +1$) was then trivial, provided that, as in Smarr and Taubes (1980), one was willing to assume that a simple perfect-fluid description is appropriate. Much of the literature on relativistic hydrodynamics considers a barotropic equation of state where the dependent variable is a function of only one quantity. Our discussion did not make this "unphysical" restriction: we have specified an equation of state that defines one function (n,e,p,T) in terms of two others, e.g., $p = p(n,T)$.

In the context of perturbation theory for general-relativistic perfect fluids, there will always be a preferred family of world lines representing the motion of typical observers in the background universe model. Thus we have adapted a coordinate system to these observers, so that b_0 equals $u_{(0)}$. Specifically, we wrote out the components of gauge-invariant variables in such an adapted coordinate system without explicitly incorporating $u_{(0)} := (u)_{|\epsilon=0}$ into the equations of perturbation theory, but rather substituting everywhere the components $u^\alpha{}_{(0)} = \delta^\alpha{}_0$ of $u_{(0)}$. At first sight, in this viewpoint, it would be concluded that the equations are not tensor equations, because the components fail to transform according to the usual law under coordinate transformations. However, the "nontensorial" nature of the theory can only be attributed to a "failure" to explicitly incorporate the extra geometrical object $u_{(0)}$ into the gauge-invariant variables. When this is done, perturbation theory will have a tensorial character, but the resulting equations will be rather more complex than those proposed here; thus they are not written here. However, there is no question that tightening a theory with further geometrical niceties always has merits.

Of course, it would be of both practical and conceptual interest to extend the present approach to situations where preferred vector fields or preferred bases of vector fields no longer exist, for example, to cases of the most general matter-associated perturbations away from a flat space-time. To the best of our knowledge, even this standard problem must still be regarded as open. When the matter is not a perfect fluid, we may wish to study perturbations using a kinetic-theory description. Accordingly, it would also seem natural to find some kinetic-theory analogs of the techniques developed in this paper and then to apply these generalized techniques to the investigation of various problems related to the Einstein–Boltzmann coupled system of equations (Banach and Piekarski, 1994a,b; Banach and Makaruk, 1995). Since Boltzmann's equation involves infinitely many degrees of freedom, it remains to be seen how useful these modified techniques will prove there. While perhaps some deeper theory of gauge-invariant perturbations would not imply the resulting physical effects were large, it nevertheless seems an important question of principle. Although a complete set of basic gauge-invariant variables corresponding to the conditions of general-relativistic kinetic theory has not yet been explicitly specified, we have already indicated how the calculation can be initiated in the case when the pressure vanishes in the background (Banach and Piekarski, 1994a).

In this paper we have considered perturbation theory on the basis of Einstein's general theory of relativity. No purported inconsistency between experiment and Einstein's laws of gravity has ever surmounted the test of time. However, our treatment is sufficiently flexible and broadly based to cover also other theories of gravity, such as, e.g., the theory of Brans and Dicke (1961) or Sławianowski's (1994) nonmetric theory. This is because any viable theory of gravity contains a gauge freedom corresponding to the group of diffeomorphisms of space-time. In the linear approximation, this again implies that two perturbations represent the same physical perturbation if (and only if) they differ by the action of an "infinitesimal diffeomorphism" on the background solution.

We hope to discuss all these modifications and generalizations in the future. Also deferred to a companion paper (Banach and Piekarski, 1996) are (i) the derivation of a closed set of differential equations governing the evolution of $D := \{\chi, \Gamma, \Omega, \Omega^r, \Delta, \Delta^{rs}, S^{ijrs}\}$ and (ii) the physical interpretation of the basic gauge-invariant quantities.

APPENDIX. FURTHER PROPERTIES OF THE BASIC VARIABLES

In this appendix, we will prove that any complicated gauge-invariant object can be expressed in terms of a complete set of basic gauge-invariant variables. Precisely speaking, our purpose here is to show that the appropriate combinations and differentiations of equations (5.21) give all these objects directly from the basic variables χ, Γ, Ω, Ω^r, Δ, Δ^{rs}, S^{ijrs} through purely algebraic operations.

In the present approach, when we speak of the complicated gauge-invariant quantity $J_h(x)$, we always have in mind the following construction: To within the ambiguity concerning the precise definition of a suitably prolonged inner product space $\mathcal{W}_x$ (a simple example of the operation of prolongation of the original structures and a hint of what to expect may be obtained from considerations of Section 5), let $\mathcal{H}_x$ be the set of $h \in \mathcal{W}_x$ such that the scalar product $\langle h,\mathcal{L}_v C_0\rangle_x$ vanishes for all vector fields v on X. Consider the equivalence class of P, denoted [P], and let h: $X \Rightarrow \mathcal{H} := \cup_{x\in X} \mathcal{H}_x$ be a cross section of $\mathcal{H}$. The gauge-invariant quantity $J_h(x)$ will then coincide with $\langle h,\mathrm{P}'\rangle_x$, where P′ is an arbitrary member of [P]; thus

$$J_h(x) := \langle h,\mathrm{P}'\rangle_x, \qquad \mathrm{P}' \in [\mathrm{P}] \tag{A.1}$$

Clearly, the value of $J_h(x)$ is completely independent of the choice of $\mathrm{P}' \in [\mathrm{P}]$ and this conclusion holds for each $x \in X$. As a result, $\langle h,\cdot\rangle_x$ defines a mapping from $\mathcal{P}/\mathcal{P}_0$ into R. Together with the identification rule $[\mathrm{P}] \Leftrightarrow$ D of Section 6, if we let D denote the system of real-valued functions χ, Γ, Ω, Ω^r, Δ, Δ^{rs}, S^{ijrs} on X, the typical expression of this observation is

$$J_h(x) = \mathrm{F}_h(x,\mathrm{D}), \qquad x \in X \tag{A.2}$$

where $\mathrm{F}_h(x,\cdot)$ denotes a functional, that is, a function whose arguments are functions D on X.

But our constructions are local, since the original variables $Q^{\alpha\beta}$, U^α, M, K are allowed to enter the definition of $\mathrm{P}(x) \in \mathcal{W}_x$ only through $Q^{\alpha\beta}(x)$, $U^\alpha(x)$, $M(x)$, $K(x)$, and their space-time derivatives of order $\leq r$; here r is an integer whose precise value depends on the choice of $J_h(x)$. Moreover, all the definitions and all the results of this paper, especially the theorem of Section 6, remain valid when we replace X by any open set $\vartheta \subset X$ in the statements and proofs. Hence we obtain

$$J_h(x) = F_h(x, D_{|\vartheta}), \qquad x \in \vartheta \tag{A.3}$$

where $D_{|\vartheta}$ is the restriction of D to ϑ. But this equation holds for all open subsets of X, however small. Thus a standard argument yields the conclusion that the value of $J_h(x)$ is uniquely specified by giving the germs of functions D at $x \in X$ [see Choquet-Bruhat *et al.* (1989) for the definition of a germ]. With this additional conclusion, equation (A.3) is to be replaced by

$$J_h(x) = F_h(x, D_x), \qquad x \in X \tag{A.4}$$

where the symbol D_x represents germs of functions D at x.

On the other hand, if we relate P(x) to the original, gauge-dependent quantities, we get from $\langle h, P\rangle_x$ an explicit expression for $J_h(x)$ which tells us that $J_h(x)$ is a linear combination of $Q^{\alpha\beta}(x),\ldots,K(x)$ and their space-time derivatives of order $\leq r$. Then in this combination we substitute the formulas that result from using the definitions (5.21) of D, so obtaining a concrete form of equation (A.4). In this way, we are able to show that the following conclusion holds: The appropriate combinations and differentiations of equations (5.21) give $J_h(x)$ directly from the set D of basic variables through purely algebraic operations once r and h in $J_h(x)$ have been specified.

Illustrations of the above observation are given in a companion paper (Banach and Piekarski, 1996).

REFERENCES

Banach, Z., and Makaruk, H. (1995). *International Journal of Theoretical Physics*, **34**, 761–777.

Banach, Z., and Piekarski, S. (1989). *Journal of Mathematical Physics*, **30**, 1804–1815.

Banach, Z., and Piekarski, S. (1994a). *Journal of Mathematical Physics*, **35**, 4809–4831, 5885–5907.

Banach, Z., and Piekarski, S. (1994b). *Journal of Statistical Physics*, **76**, 1415–1437.

Banach, Z., and Piekarski, S. (1996). Gauge-invariant perfect-fluid Robertson–Walker perturbations, *International Journal of Theoretical Physics*, this issue.

Bardeen, J. M. (1980). *Physical Review D*, **22**, 1882–1905.

Brans, C., and Dicke, R. H. (1961). *Physical Review*, **124**, 925–935.

Choquet-Bruhat, Y., DeWitt-Morette, C., and Dillard-Bleick, M. (1989). *Analysis, Manifolds, and Physics. Part I: Basics*, North-Holland, Amsterdam.

D'Eath, P. D. (1976). *Annals of Physics*, **98**, 237–263.

Ehlers, J. (1973). In *Relativity, Astrophysics, and Cosmology*, W. Israel, ed., Reidel, Boston, Volume 38, pp. 1–125.

Ellis, G. F. R., and Bruni, M. (1989). *Physical Review D*, **40**, 1804–1818.

Emch, G. G. (1972). *Algebraic Methods in Statistical Mechanics and Quantum Field Theory*, Wiley-Interscience, New York.

Fischer, A. E., Marsden, J. E., and Moncrief, V. (1980). In *Essays in General Relativity*, F. J. Tipler, ed., Academic Press, New York, pp. 79–96.

Gerlach, U. H., and Sengupta, U. K. (1978). *Physical Review D*, **18**, 1789–1797.

Greub, W. (1975). *Linear Algebra*, Springer-Verlag, New York.

Hawking, S. W. (1966). *Astrophysical Journal*, **145**, 544–554.
Lifshitz, E. M. (1946). *Zhurnal Eksperimental'noi i Theoreticheskoi Fiziki*, **16**, 587.
Lifshitz, E. M., and Khalatnikov, I. M. (1963). *Advances in Physics*, **12**, 185–249.
Meyer, J., and Schröter, J. (1981). *Archives for Rational Mechanics and Analysis*, **76**, 193–246.
Misner, C. W., Thorne, K. S., and Wheeler, J. A. (1973). *Gravitation*, Freeman, San Francisco.
Peebles, P. J. E. (1993). *Principles of Physical Cosmology*, Princeton University Press, Princeton, New Jersey.
Sachs, R. K. (1964). In *Relativity, Groups, and Topology*, C. DeWitt and B. S. DeWitt, eds., Gordon and Breach, New York, pp. 520–562.
Sławianowski, J. J. (1994). *Archives of Mechanics*, **46**, 375–397.
Smarr, L., and Taubes, C. (1980). In *Essays in General Relativity*, F. J. Tipler, ed., Academic Press, New York, pp. 157–183.
Szegö, G. (1939). *Orthogonal Polynomials*, American Mathematical Society, Providence, Rhode Island.
Truesdell, C., and Toupin, R. (1960). In *Handbuch der Physik*, S. Flügge, ed., Springer-Verlag, Berlin, pp. 226–793.
Wald, R. (1984). *General Relativity*, University of Chicago Press, Chicago.