TOPICAL REVIEW

A concise introduction to perturbation theory in cosmology

Karim A Malik\textsuperscript{1} and David R Matravers\textsuperscript{2}

\textsuperscript{1} Astronomy Unit, School of Mathematical Sciences, Queen Mary University of London, Mile End Road, London, E1 4NS, UK
\textsuperscript{2} Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 2EG, UK

Received 29 April 2008
Published 16 September 2008
Online at stacks.iop.org/CQG/25/193001

Abstract
We give a concise, self-contained introduction to perturbation theory in cosmology at linear and second orders, striking a balance between mathematical rigour and usability. In particular, we discuss gauge issues and the active and passive approaches to calculating gauge transformations. We also construct gauge-invariant variables, including the second-order tensor perturbation on uniform curvature hypersurfaces.

PACS numbers: 98.80.Jk, 98.80.Cq

1. Introduction

Cosmological perturbation theory has recently enjoyed renewed interest. Linear or first-order theory is still a very active field of research, though the focus has moved on to higher order and even fully nonlinear theory. This is to a large extent due to the availability of much improved data sets: whereas previously linear theory was sufficient and the power spectrum was the observable of choice, now the quality and quantity of the data are such, that higher-order observables, as for example the bispectrum, can be compared with the theoretical predictions. These new data sets come from observations of the cosmic microwave background (CMB) on the one hand, such as the one already in progress by \textit{WMAP} and in the near future also by \textit{PLANCK}. But also from 21 cm surveys on the other hand, mapping the anisotropies in neutral hydrogen, such as \textit{LOFAR}, now under construction and \textit{SKA}, which is currently in its design phase.

Einstein’s theory of general relativity (GR) is highly nonlinear; it is therefore difficult to deal with in all but the simplest situations using the full theory. Fortunately for cosmologists the universe appears to be homogeneous and isotropic to a remarkable degree so the Friedmann–Robertson–Walker (FRW) metric is adequate for many purposes. For instance, once known local features are removed, the CMB is isotropic to an accuracy of $\delta T/T$ of $10^{-5}$. However if we want greater resolution or more detail then the approximation has to take into account
anisotropy and inhomogeneity. At present this cannot be done in full generality since we do not have the appropriate exact solutions to Einstein’s equations. This is not surprising, given their highly nonlinear nature. To deal with this problem cosmologists have resorted to perturbation methods, which have proved effective in other areas of physics. Previous relevant works on perturbation theory in cosmology at a linear order include [1–7], and at second order [8–13]. Beyond linear order the literature tends to be very technical and difficult for the non-specialist to follow. In this paper, we aim to strike a balance between mathematical rigour using the language and tools of differential geometry, and usability and applicability to the problems of theoretical astrophysics and cosmology.

The essential idea behind perturbation theory is very simple, and best illustrated by an example for which we choose the metric tensor in standard cosmology. We assume that we can approximate the full metric \( g_{\mu\nu} \) of the universe by an expansion

\[
g_{\mu\nu} = g_{\mu\nu}^{(0)} + g_{\mu\nu}^{(1)} + \frac{1}{2} g_{\mu\nu}^{(2)} + \cdots.
\]  

(1.1)

The metric \( g_{\mu\nu}^{(0)} \), called the background, is the FRW metric with appropriate spatial curvature, i.e. \( K = 0, 1, -1 \) according to the assumptions made about the universe. The remaining terms are the perturbations of the background. The first-order part is given by

\[
g_{\mu\nu} - g_{\mu\nu}^{(0)} = g_{\mu\nu}^{(1)},
\]  

(1.2)

where the remaining terms are assumed to be negligible compared to \( g_{\mu\nu}^{(0)} \), and are neglected at first order. In a similar way the higher-order perturbations can be identified. This can be described simply if we assume that the series can be written as

\[
g_{\mu\nu} = g_{\mu\nu}^{(0)} + \epsilon \tilde{g}_{\mu\nu}^{(1)} + \epsilon^2 \tilde{g}_{\mu\nu}^{(2)} + \cdots,
\]  

(1.3)

where the quantities with tildes have absolute magnitudes of order unity, and we assume that \( \epsilon \ll 1 \). To zeroth order we have \( g_{\mu\nu} = g_{\mu\nu}^{(0)} \) and at first order

\[
g_{\mu\nu} = g_{\mu\nu}^{(0)} + \epsilon \tilde{g}_{\mu\nu}^{(1)},
\]  

(1.4)

and so on using the fact that at each order the higher-order terms can be ignored. In practice it is often a nuisance to introduce the parameter \( \epsilon \), where appropriate we will use the form (1.1). Issues of convergence can be removed by working within a small enough neighbourhood of the background.

Having set up the approximation (1.1), we have to substitute it into the Einstein equations

\[
G_{\mu\nu} + \Lambda g_{\mu\nu} = \kappa T_{\mu\nu}
\]  

(1.5)

to obtain approximate solutions at the required order of approximation for the application we have in mind. This is more difficult than one might imagine. First, perturbations of the metric imply perturbations of the energy–momentum tensor, but more importantly, calculation of the connection coefficients and the Ricci tensor involves raising and lowering indices and involves terms of different orders. At zero and first order this is not a problem, but at higher orders it makes the calculations much more complicated and so the choice of coordinates or the form of the metric can be important. Already at second order we have ‘proper’ second-order terms and terms quadratic in the first-order quantities.

Another problem arising in cosmological perturbation theory is the presence of spurious coordinate artefacts or gauge modes in the calculations. Although GR is covariant, i.e. manifestly coordinate choice independent, splitting variables into a background part and a perturbation is not a covariant procedure, and therefore introduces this gauge dependence. Prior to 1980 the gauge modes were handled on a case-by-case basis, but when Bardeen in [5] resolved the issue and provided a systematic procedure for eliminating the gauge freedom at first order. Although it is sometimes argued that the covariant approach [14] avoids the issue
of gauge choice, it corresponds to the comoving gauge which is made explicit by the inclusion of the velocity field [15]. Below we will address the gauge issue in some detail and explain how it can be resolved at first and at second order.

The paper is organised as follows. In the following section, we introduce perturbation theory using notation and concepts from differential geometry. In particular, we discuss the definition of perturbations and how perturbations change under small coordinate changes. In section 3 we apply the concepts and results of section 2. We discuss the construction of gauge-invariant variables at first and second orders. Amongst the examples discussed is the second-order tensor perturbation and how it can be rendered gauge invariant. We discuss our results in section 4. We finish this paper with an appendix in which we describe the relevant concepts from differential geometry used in section 2.

We predominantly use conformal time, \( \eta \), related to coordinate time \( t \) by \( dt = a \, d\eta \), where \( a \) is the scale factor. Derivatives with respect to conformal time are denoted by a dash. Greek indices \( \mu, \nu, \lambda \) run from 0, ..., 3, upper case Latin indices \( A, B, C \) run from 0, ..., 4, while lower case Latin indices \( i, j, k \) run from 1, ..., 3.

2. Perturbation theory

In this section we introduce perturbation theory using differential geometry. Though focusing on cosmology, we keep the discussion general. After giving a definition of perturbations we introduce and define the concept of gauge and study how perturbations change under gauge transformations. The relevant definitions from differential geometry are discussed in the appendix.

2.1. Cosmological perturbation theory: perturbations of spacetime

The application of perturbation methods in spacetime brings in a new problem, since among the physical quantities to be perturbed is the spacetime itself. Also, because the results will be used in relativistic cosmology, the theory and results must be covariant. These requirements lead us to base our discussion on the explicitly coordinate-independent description of Sachs [1], Stewart and Walker [2] and Stewart [3]. In this case by coordinate independent we mean that the description does not require coordinates and so is intrinsically covariant. We can and do introduce coordinates to do calculations and simplify the exposition. An alternative and widely used, but coordinate dependent, description is given in [4]. As one would expect the results are the same although, in our view, it is easier to understand and see the source of the final equations in the description of [3]. We will however describe both procedures and show the connection between the two approaches.

We now follow Stewart [3] closely and consider a one-parameter family of 4-manifolds \( \mathcal{M}_\epsilon \) embedded in a 5-manifold \( \mathcal{N} \). Each manifold in the family represents a perturbed spacetime with the base or unperturbed spacetime manifold represented by \( \mathcal{M}_0 \). We define a point identification map \( P_\epsilon : \mathcal{M}_0 \to \mathcal{M}_\epsilon \) which identifies points in the unperturbed manifolds with points in the perturbed manifold. This correspondence specifies a vector field \( X \) upon \( \mathcal{N} \). This field is transverse to \( \mathcal{M}_\epsilon \) at all points. The points which lie on the same integral curve \( \gamma \) of \( X \) are to be regarded as the same point, see figure 1. This can be expressed in terms of coordinates. Choose coordinates \( x^a \) on \( \mathcal{M}_0 \) and extend them to \( \mathcal{N} \) by requiring that \( x^a = \) constant along each of the curves \( \gamma \). This induces coordinates \( \{x^A = (x^a, \epsilon)\} \) with \( A = 0, 1, \ldots, 4 \) \( \mu, \nu, \ldots = 0, 1, \ldots, 3 \) on \( \mathcal{N} \). We parametrize the curves \( \gamma \) by \( \epsilon \) and so \( dx^A / d\epsilon = X^A \) and we choose the scaling of \( \epsilon \) such that

\[
\phi_\epsilon : \mathcal{M}_0 \to \mathcal{M}_\epsilon. \tag{2.1}
\]
In this way the vector field $X$ generates a one-to-one, invertible, differentiable mapping between $\mathcal{M}_0$ and $\mathcal{M}_\epsilon$, i.e. a one-parameter group of diffeomorphisms and it follows that $\phi_\epsilon \phi_{\epsilon^{-1}} = \phi_{\epsilon^{-1}} \phi_\epsilon$.

In particular the inverse map from $\mathcal{M}_\epsilon$ to $\mathcal{M}_0$ will be denoted, in an obvious notation, by $\phi_{\epsilon^{-1}}$ and the identity map is given by $\phi_{0}$.

Given a geometric quantity $T$ defined on $\mathcal{N}$ the simplest way to produce a perturbation expansion of $T$ is to expand it as a Taylor series along $\gamma$. This yields a covariant power series for $T$ along the curve. To first order the series has the form

$$\phi_\epsilon^* T_\epsilon = T_0 + \epsilon (\mathcal{L}_X T)|_0 + O(\epsilon^2),$$

(2.2)

where $\phi_\epsilon^*$ is used to indicate that the quantity is the pullback, i.e. it is $T_\epsilon$ evaluated at the point where $\epsilon = 0$. Lie derivatives are used instead of partial derivatives so that the series is covariant. For reasons that will become obvious later it is convenient that the series is pulled back to $\mathcal{M}_0$ (see equation (A.20)). At higher orders the Taylor expansion is given by

$$\phi_\epsilon^* T_\epsilon = T_0 + \sum_{j=1}^{\infty} \frac{\epsilon^j}{j!} (\mathcal{L}_X^j T)|_0,$$

(2.3)

where we note again that $\phi_\epsilon^* T_\epsilon$ is evaluated on $\mathcal{M}_0$. The expansion automatically provides the covariant perturbation expansion we want. Each term in the series is proportional to a power of $\epsilon$. The first term $T_0$ is proportional to $\epsilon^0$, the background value, the next term $\epsilon (\mathcal{L}_X T)|_0$ is proportional to $\epsilon$ to the first order and so on, and the $n$th order term is given by $\frac{\epsilon^n}{n!} (\mathcal{L}_X^n T)|_0$.

The expansion equation (2.3) can be written in a compact and useful form using the exponential operator

$$\phi_\epsilon^* T_\epsilon = (e^{\epsilon \mathcal{L}_X}) T|_0.$$  

(2.4)

Here $\phi_\epsilon^* T_\epsilon$ is the perturbed value of $T$ pulled back to $\mathcal{M}_0$ and so the perturbed value of $T$ is given by

$$\phi_\epsilon^* \delta T_\epsilon = \phi_\epsilon^* T_\epsilon - T_0,$$

where we note that we could not have done the subtraction if we had not pulled $T_\epsilon$ back to $\mathcal{M}_0$. In an alternative notation, commonly used in the literature, we include the $\epsilon$ with the $T$ and write

$$T = T_0 + \delta T,$$

(2.5)
Figure 2. On the left panel, the passive view: the point $p$ on the manifold $M_0$ is mapped to two different points $q$ and $u$ on $M_\epsilon$ depending on the choice of gauge, corresponding to the choice of vector field, we make. On the right panel, the active view: the points $p$ and $q$ on $M_0$ both map to the point $u$ on $M_\epsilon$. Again the choice of gauge determines the mapping. The vector fields generate the gauge choice. A change in gauge from $X^A$ to $Y^A$ produces a gauge transformation.

where

$$\delta T = T_1 + \frac{1}{2}T_2 + \frac{1}{3!}T_3 + \cdots ,$$

(2.6)

with $T_n = \epsilon^n (\ell^n T)_0$.

In [4] the approach is as follows. On a single spacetime manifold $\mathcal{M}$ with coordinates $x^\mu$ define a background model by assigning to all geometric fields $Q$ a fixed background value $Q^0$, which is not itself a geometric quantity, at each point on the manifold. While the fields $Q$ may transform as scalar, vector or tensor fields we require that the $Q^0$ be fixed functions of the coordinates. Under a coordinate transformation the $Q^0$ will have the same functional dependence on the new coordinates as they had on the old ones. A perturbation is then given by

$$\delta Q = Q - Q^0.$$  

(2.7)

To relate the two approaches we can think of the $Q^0$ quantity playing the role of a quantity defined on $M_0$ in the Stewart description and the coordinate change corresponding to a change of coordinates on $\mathcal{M}_\epsilon$. But it is important to note that in the approach of [4] only one manifold is necessary. The Stewart approach [3] avoids the need for the quantity $Q^0$, which is not covariant and gives a simple diagrammatic representation at the price of having to introduce the abstract five-dimensional manifold $\mathcal{N}$. However, note that it splits into a background and a perturbation which in general is not covariant. This split is common to both approaches and it gives rise to the gauge dependence.

2.2. Gauge transformations

Gauge is arguably the most over-used word in mathematics and physics. Sometimes the meanings are related but often they are not and it is a waste of time trying to relate them. To avoid confusion we recommend that the word ‘gauge’ as used here is interpreted as defined and not related to other uses of the word. The choice of correspondence between points on $M_0$ with those on $M_\epsilon$ or, equivalently, the choice of a vector field $X$ is a gauge choice. The vector field $X$ is called the generator of the gauge.

Let us now turn to defining gauge dependence in a clearer way. Consider a point $p$ in $M_0$ and the generators $X$ and $Y$ corresponding to two different gauge choices (see figure 2). The choice $X$ will identify point $p$ on $M_0$ with a point $q$ on $M_\epsilon$ and will assign to $q$ the same $x^\mu$.
coordinates as at point $p$. On the other hand, the gauge choice $Y$ will identify $p$ with a different point $u$ on $M_{\epsilon}$ assigning in its turn the coordinates of $p$ to $u$. Clearly, the choice of gauge induces a coordinate change (a gauge transformation) on $M_{\epsilon}$. This interpretation is called the passive view [4, 8].

For the active view, we choose a point $u$ on $M_{\epsilon}$ and find the point $p$ on $M_0$ which maps to $u$ under the gauge choice $X$ and the point $q$, also on $M_0$, which maps to $u$ under the gauge choice $Y$, see figure 3. The gauge transformation this time is defined on $M_0$ and takes the coordinates of $q$ to those of $p$ in one of the two choices of gauge.

In summary as we shall explain in more detail below, in the active approach the transformation of the perturbed quantities is evaluated at the same coordinate point, whereas in the passive approach the transformation is taken at the same physical point.

In the passive approach of [4] the role of the background manifold is played by the background quantities $Q^0$ and the coordinate transformation corresponding to the gauge choice only affects the geometric quantities $Q$. The perturbation is the difference between $Q^0$ and $Q$ so only half the quantities determining the perturbation are transformed by the gauge transformation.

The gauge dependence in perturbation theory stems from the fact that we separate quantities into a background and a perturbed part, a operation not covariant in general, which introduces additional, unphysical degrees of freedom. However, as shown in section 3, by choosing and combining suitable matter and metric variables the gauge dependences can be made to cancel out (the quantities so constructed will not change under a gauge transformation). This process is equivalent to choosing suitable physical hypersurfaces, say comoving or of uniform curvature.

2.2.1. Active point of view. To take the argument further we will now focus on the active interpretation of the gauge transformation. Corresponding to the gauge choice $X$, i.e. the choice

\footnote{One should not confuse gauge independence in perturbation theory with what is called gauge choice in general relativity which arises from coordinate invariance: the Bianchi identities introduce four additional degrees of freedom into the Einstein equations. This allows us to always choose four free functions in the metric, i.e. four particular coordinate functions, that might simplify the problem under consideration if suitably chosen.}
of the vector field $X$ transverse to $\mathcal{M}_0$ we have a diffeomorphism $\phi_\epsilon$ where $\phi_\epsilon : \mathcal{M}_0 \rightarrow \mathcal{M}_\epsilon$ and corresponding to the vector field $Y$ we have a diffeomorphism $\psi_\epsilon : \mathcal{M}_0 \rightarrow \mathcal{M}_\epsilon$. For all $\epsilon$ these two vector fields induce a diffeomorphism (gauge transformation) $\Phi_\epsilon$ on $\mathcal{M}_0$ given by, see figure 3,

$$\Phi_\epsilon : \mathcal{M}_0 \rightarrow \mathcal{M}_0,$$  \hspace{1cm} (2.8)

where $\Phi_\epsilon$ is made up of two parts—a map $\psi_\epsilon$ from $\mathcal{M}_0$ to $\mathcal{M}_\epsilon$ and a map $\phi_{-\epsilon}$ from $\mathcal{M}_\epsilon$ to $\mathcal{M}_0$, i.e.,

$$\Phi_\epsilon := \phi_{-\epsilon} \circ \psi_\epsilon.$$  \hspace{1cm} (2.9)

Under this gauge transformation the transformation of a geometric quantity $T$ is given by (see equations (A.16) and (A.18))

$$\Phi_{\epsilon} T = (\phi_{-\epsilon} \circ \psi_\epsilon)_* T$$

$$= \psi_{-\epsilon} \circ \phi_\epsilon T$$

$$= e^{(\epsilon \xi_1)} e^{(-\epsilon \xi_2)} T,$$  \hspace{1cm} (2.10)

(2.11)

(2.12)

where we have used the fact that the pull-backs of the transformations induced by the gauge choices can be written as Taylor series in terms of the exponential notation as

$$\phi_{\epsilon} T = e^{\epsilon \xi_1} T,$$

$$\psi_{\epsilon} T = e^{\epsilon \xi_2} T$$

(2.13)

(2.14)

(for more details again see equations (A.16)–(A.18)). Also note that the $T$ here have to be evaluated on $\mathcal{M}_0$ but putting $T_0$ would be confusing.

Now we invoke the Baker–Campbell–Haussdorf formula [16] which enables us to write $\Phi_{\epsilon} T$ in the following form:

$$\Phi_{\epsilon} T = \exp \left( \sum_{n=1}^{\infty} \frac{\epsilon^n}{n!} \xi_n \right) T,$$  \hspace{1cm} (2.15)

where

$$\xi_1 = Y - X, \quad \xi_2 = [X, Y], \quad \text{and} \quad \xi_3 = \frac{1}{2} [X + Y, [X, Y]],$$

and $X$ and $Y$ are the gauge generators, i.e. the vectors which determine the gauge choices. Explicitly, the first few terms of the gauge transformation equation (2.15) are

$$\Phi_{\epsilon} T = T|_0 + \epsilon \xi_1 T|_0 + \frac{\epsilon^2}{2} \left( \xi_2 + \xi_1^2 \right) T|_0 + \frac{\epsilon^3}{3!} \left( \xi_3 + \frac{3}{2} \left[ \xi_1, \xi_2 \right] + \xi_1^3 \right) T|_0 + O(\epsilon^4),$$

(2.16)

where we indicate that $T$ has to be evaluated on the manifold $\mathcal{M}_0$ by the notation $T|_0$.

If we now use equation (2.5) to introduce the results of Taylor expanding $T$ into equation (2.16) we obtain

$$\bar{T}_0 = T_0, \hspace{1cm}$$

$$\bar{T}_1 = T_1 + \xi_1 T_0,$$

$$\bar{T}_2 = T_2 + \xi_2 T_0 + \xi_1^2 T_0 + 2 \xi_1 T_1,$$

where $\xi^k$ is the vector field generating the transformation and $\xi^\mu \equiv \epsilon \xi^\mu_1 + \frac{1}{2} \epsilon^2 \xi^\mu_2 + O(\epsilon^3).$
Similarly, the map \( (U, x) \xrightarrow{\Phi} (U', \tilde{x}) \) (see figure 1) under an infinitesimal transformation generated by \( \epsilon \xi^\mu \). In the active view, this transformation takes the point \( p \) with coordinates \( x^\mu(p) \) to the point \( q = \Phi_\epsilon(p) \) with coordinates \( x^\mu(q) \). Note that in the active view it is the points that change. Applying the map (2.15) it follows

\[
x^\mu(q) = e^{\frac{\epsilon}{2} \xi^\mu(x(p))}, \tag{2.18}
\]

where we have used the fact that when acting on scalars \( \xi^\mu = \xi^\mu \frac{\partial}{\partial x^\mu} \) and the partial derivatives are evaluated at \( p \).\(^4\) The left-hand side and the right-hand side of equation (2.18) are evaluated at different points. Equation (2.18) can then be expanded up to second order as

\[
x^\mu(q) = x^\mu(p) + \epsilon \xi^\mu_1(p) + \frac{1}{2} \epsilon^2 \left[ \xi^\mu_{1\nu}(x(p)) \xi^\nu_1(p) + \xi^\nu_2(p) \right]. \tag{2.19}
\]

Note that we do not need equation (2.19) to calculate how perturbations change under a gauge transformation in the active approach, it simply tells us how the coordinates of the points \( p \) and \( q \) are related in this approach.

2.2.2. Passive point of view. In the passive approach we specify the relation between two coordinate systems directly, and then calculate the change in the metric and matter variables when changing from one system to the other. As long as the two coordinate systems are related through a small perturbation, the functional form relating them is quite arbitrary. However, in order to make contact with the active approach, discussed above, we take equation (2.19) as our starting point.

Note that all quantities in the passive approach are evaluated at the same physical point. To take the passive approach further, we therefore need to rewrite the left-hand side and the right-hand side of equation (2.19), since they are evaluated at two different coordinate points, as described above (see also figure 2). We choose \( p \) and \( q \) to be points such that the coordinates of \( q \) in the new coordinates are the same as the coordinates of \( p \) in the old coordinates, i.e. \( \tilde{x}^\mu(q) = x^\mu(p) \), then use equation (2.19) to derive

\[
\tilde{x}^\mu(q) = x^\mu(p) - \epsilon \xi^\mu_1(p) - \epsilon \frac{1}{2} \xi^\mu_{1\nu}(x(p)) \xi^\nu_1(p) - \xi^\nu_2(p). \tag{2.20}
\]

Using the first terms of equation (2.19) we have

\[
x^\mu(q) = x^\mu(p) + \epsilon \xi^\mu_1(p), \tag{2.21}
\]

to get a Taylor expansion for \( \xi^\mu_1 \),

\[
\xi^\mu_1(p) = \xi^\mu_1(x^\mu(q) - \epsilon \xi^\mu_1(p)) = \xi^\mu_1(q) - \epsilon \xi^\mu_1(q) \xi^\nu_1(q). \tag{2.22}
\]

where in the very last term we have replaced \( \xi^\nu_1(p) \) by \( \xi^\nu_1(q) \), the correction being of third order. Substituting equation (2.22) into equation (2.20) finally gives the desired result, namely a relation between the ‘old’ (untilded) and the ‘new’ (tilde) coordinate systems,

\[
\tilde{x}^\mu(q) = x^\mu(q) - \epsilon \xi^\mu_1(q) - \epsilon \frac{1}{2} \left[ \xi^\nu_2(q) - \xi^\nu_1(q) \xi^\nu_1(q) \right]. \tag{2.23}
\]

all evaluated at the same point \( q \).

\(^4\) Note that if we had used a minus sign in the exponent of equation (2.18) the signs in the equation relating the coordinates in the passive approach, equation (2.23), would conform to those usually found in the literature.
3. Applications

As an application and illustration of the above we now derive the transformation behaviour under gauge transformations of some quantities at first and second orders. We start at first order by highlighting the two different points of view in how the vector fields inducing the coordinate change affect the perturbations, as detailed in section 2.2.

Before studying the transformation behaviour of the perturbations, we define and relate them to their respective backgrounds in the following. As our first example we choose a 4-scalar, we use here the energy density \( \rho \), which can be expanded up to second order using equation (2.5),

\[
\rho = \rho_0 + \delta \rho_1 + \frac{1}{2} \delta \rho_2,
\]

where we already split \( \delta \rho \) into its first- and second-order parts according to equation (2.6), the subscripts denoting the order of the perturbations.

Our second example is given by the metric tensor \( g_{\mu \nu} \), as outlined in equation (1.1). In particular, using equation (2.5), the complete Friedmann–Robertson–Walker metric tensor, up to and including second-order perturbations, can be written as

\[
\begin{align*}
g_{00} &= -a^2(1 + 2\phi_1 + \phi_2), \\
g_{0i} &= a^2(B_{1i} + \frac{1}{2} B_2), \\
g_{ij} &= a^2[\delta_{ij} + 2C_{1ij} + C_{2ij}],
\end{align*}
\]

where \( a \) is a flat \((K = 0)\) background.

The first- and second-order perturbations \( B_{1i} \) and \( C_{1ij}, B_{2i} \) and \( C_{2ij} \), can be further split according to equations (3.5) and (3.6) into scalar, vector and tensor parts (defined according to their transformation behaviour on spatial 3-hypersurfaces),

\[
\begin{align*}
B_i &= B_i - S_i, \\
C_{ij} &= -\psi \gamma_{ij} + E_{ij} + F_{(ij)} + \frac{1}{2} h_{ij},
\end{align*}
\]

where the vector parts, \( S_i \) and \( F_i \), are divergence free, and the tensor part, \( h_{ij} \), is divergence free and traceless, i.e.,

\[
S^k_k = 0, \quad F^k_k = 0, \quad h^{ik}_k = 0, \quad h^k_k = 0.
\]

The order of the perturbations in equations (3.5)–(3.7) has been omitted in the above for ease of presentation. Note that \( \psi \) is the curvature perturbation, describing the intrinsic scalar curvature of spatial hypersurfaces. Furthermore \( \phi \) is the lapse function, \( h_{ij} \) is the tensor perturbation describing the gravitational wave content, \( B \) and \( E \) describe the scalar shear and \( S_i \) and \( F_i \) the vector part of the shear.

Here and in the following we assume a flat background without loss of generality, just simplifying our calculations and allowing us to use partial derivatives in expressions such as equation (3.6).

The perturbations are decomposed into scalar, vector and tensor parts since at linear order the governing equations for the different types decouple. This is however no longer the case at higher orders, and indeed we already see from the gauge transformations and the definitions of gauge-invariant variables at second order that, e.g., the energy density (a scalar quantity) on flat hypersurfaces now also contains first-order vector and tensor parts, see equation (3.53).

Finally, we should point out that the decomposition of the metric tensor in equation (3.2) is not unique. This is already evident in the temporal part of the metric tensor, where the
lapse function \(\phi\) is here simply expanded in a power series, \(\phi = \phi_1 + \frac{1}{2}\phi_2 + \cdots\). Alternatively we could have expanded \(\exp(\phi)\) in a power series, this obviously does not affect the physics. More importantly, also the decomposition of the spatial part of the metric tensor, that is equation (3.6) is not unique. Indeed, other decompositions are in use and can be just as useful or better, depending on the circumstances and the application intended. For example it can be useful instead of expanding \(\psi\) in equation (3.6) directly in a power series, to expand \(e^\psi\) (see, e.g., [17], and for a relation of the two expansions [18]).

3.1. Passive point of view

The passive point of view is very popular at first order, see, e.g., the original paper by Bardeen [5], the review by Kodama and Sasaki [6] and the one by Mukhanov, Feldman and Brandenberger [4].

The starting point in the passive approach is to identify an invariant quantity, that allows us to relate quantities to be evaluated in the two coordinate systems. We denote the two coordinate systems by \(\tilde{x}^\mu\) and \(x^\mu\) system, and their relation is given by equation (2.23). We choose as an example the energy density, \(\rho\), which as a 4-scalar would not change (however, once it has been split into different orders, it will change). Another invariant is the line element \(ds^2\), which allows us to study the transformation properties of the metric tensor, by exploiting the invariance of \(ds^2\), i.e.,

\[
    ds^2 = g_{\mu\nu} d\tilde{x}^\mu d\tilde{x}^\nu = g_{\mu\nu} dx^\mu dx^\nu,
\]

which we will not pursue here, but see, e.g., [6, 19].

Turning instead to the energy density as an illustrative example, we get the transformation behaviour of the perturbation from the requirement that it has to invariant under a change of coordinate system and therefore has to be the same in the \(\tilde{x}^\mu\) and the \(x^\mu\) system, that is

\[
    \tilde{\rho}(\tilde{x}^\mu) = \rho(x^\mu).
\]

To first order, the two coordinate systems are related, using the linear part of equation (2.23), by

\[
    \tilde{x}^\mu = x^\mu - \xi_1^\mu.
\]

Before we study the transformation behaviour of the perturbations at first order, we split the generating vector \(\xi_1^\mu\) into a scalar temporal part \(\alpha_1\) and a spatial scalar and vector part, \(\beta_1\) and \(\gamma_1^i\), according to

\[
    \xi_1^\mu = (\alpha_1, \beta_1^i + \gamma_1^i),
\]

where the vector part is divergence free \(\partial_j \gamma_1^j = 0\). Then expanding the left-hand side of equation (3.9), we get neglecting terms of \(O(\epsilon^2)\),

\[
    \tilde{\rho}(\tilde{x}^\mu) = \rho(x^\mu - \xi^\mu) \\
    = \tilde{\rho}(x^\mu) - \tilde{\rho} \xi^\mu \\
    = \rho_0 + \delta \rho_1 - \rho_0 \alpha_1,
\]

and similarly expanding the right-hand side of equation (3.9), we have

\[
    \rho(x^\mu) = \rho_0(x^\mu) + \delta \rho_1(x^\mu).
\]

Finally, since by assumption \(\rho_0(x^\mu) = \rho_0(x^\mu)\), we get

\[
    \delta \rho_1 = \delta \rho_1 + \rho_0 \alpha_1.
\]

Note that all quantities are evaluated at the same physical point.
3.2. Active point of view

We now turn to the active point of view when calculating the effect of gauge transformations on perturbations. Here, as detailed in section 2.2, one actively maps the perturbed quantities from one manifold to another. The relation of the coordinate systems on the two manifolds is also induced by the map.

At first order the preference of which approach to use is a question of taste, and as pointed out above most first-order papers use passive viewpoint, but see, e.g., [20] for first-order active calculation. However, at second order we found the active viewpoint easier to implement, and it is used in many other second-order works, e.g. [9, 8, 11].

3.2.1. First order. As in the passive view section above, we start with the energy density. It follows immediately from equation (2.17) that to first order a scalar quantity such as the energy density transforms as

\[ \tilde{\delta}\rho_1 = \delta\rho_1 + \rho_0^'\alpha_1. \] (3.15)

The transformations of the first-order metric perturbations also follow from equation (2.17). We then find that the metric tensor transforms at first order as

\[ \tilde{\delta}g^{(1)}_{\mu\nu} = \delta g^{(1)}_{\mu\nu} + g^{(0)}_{\mu\lambda} \xi^{(1)}_{\lambda\nu} + g^{(0)}_{\nu\lambda} \xi^{(1)}_{\lambda\mu}. \] (3.16)

As another example we now turn to the spatial part of the metric tensor. Note that equation (2.17) gives only the transformation of the total spatial part of the metric, \( C_{ij} \). If we then ask how the components of \( C_{ij} \) transform, we have to use equation (3.7).

To get the change of the metric functions in the spatial part of the metric under a gauge transformation, we get the transformation of the spatial part of the metric \( \delta g^{(1)}_{ij} \), and hence \( C^{(1)}_{ij} \) from equation (3.16) as

\[ 2\tilde{C}_{1ij} = 2C_{1ij} + 2\mathcal{H}\alpha_1\delta_{ij} + \xi_{1i,j} + \xi_{1j,i}. \] (3.17)

where we reproduce equation (3.6) for convenience at first order,

\[ 2C_{1ij} = -2\psi_1\delta_{ij} + 2E_{1,ij} + 2F_{1(i,j)} + h_{1ij}. \] (3.18)

Taking the trace of equation (3.17) and substituting into equation (3.18) we get

\[ -3\tilde{\psi}_1 + \nabla^2\tilde{E}_1 = -3\psi_1 + \nabla^2E_1 + 3\mathcal{H}\alpha_1 + \nabla^2\beta_1. \] (3.19)

Now applying the operator \( \partial^i\partial^j \) to equation (3.17) we get a second equation relating the scalar perturbation \( \psi_1 \) and \( E_1 \),

\[ -3\nabla^2\tilde{\psi}_1 + \nabla^2\nabla^2\tilde{E}_1 = -3\nabla^2\psi_1 + \nabla^2\nabla^2E_1 + 3\mathcal{H}\nabla^2\alpha_1 + \nabla^2\nabla^2\beta_1. \] (3.20)

Taking the divergence of equation (3.17) we get

\[ 2\tilde{\nabla}^2l_{ij} = 2C_{ij} + 2\mathcal{H}\alpha_{1,i} + \nabla^2\xi_{ii} + \nabla^2\beta_{i,i}. \] (3.21)

Substituting in our results for \( \tilde{\psi}_1 \) and \( \tilde{E}_1 \) we then arrive at

\[ \nabla^2\tilde{F}_{li} = \nabla^2F_{li} + \nabla^2\gamma_{li}. \] (3.22)

We can sum up the well-known transformations of the first-order metric perturbations we have from the above, first for the scalars as (e.g. [20])

\[ \tilde{\phi}_1 = \phi_1 + \mathcal{H}\alpha_1 + \alpha'_i, \] (3.23)

\[ \tilde{\psi}_1 = \psi_1 - \mathcal{H}\alpha_1, \] (3.24)

\[ \tilde{B}_1 = B_1 - \alpha_1 + \beta'_i, \] (3.25)
\[ \tilde{E}_1 = E_1 + \beta_1, \]  
(3.26)

where \( \mathcal{H} = a'/a \), and for the vector perturbations as

\[ \tilde{S}^i_1 = S^i_1 - \gamma^i_1', \]  
(3.27)

\[ \tilde{F}^i_1 = F^i_1 + \gamma^i_1. \]  
(3.28)

The first-order tensor perturbation is found to be gauge invariant

\[ \tilde{h}_{ij} = h_{ij}, \]  
(3.29)

by substituting equations (3.23)–(3.28) into equation (3.17). This can also be understood from the Stewart–Walker lemma [2]: at first order, quantities that are identically zero in the background are manifestly gauge invariant, and there is no tensor part in the background. However, as we shall see below, this only works for quantities at the next higher order: for example, the second-order tensor perturbations will in general not be gauge invariant.

3.2.2. Constructing gauge-invariant variables at first order. To construct a gauge-invariant quantity, say the energy density on flat slices, that is hypersurfaces on which \( \tilde{\psi}_1 = 0 \), we see from equation (3.24) that gives

\[ \alpha_1 = \frac{\psi_1}{\mathcal{H}}. \]  
(3.30)

All we need to do next is to substitute equation (3.30) into equation (3.15), and get a gauge invariant in the sense of being independent of gauge artefacts, for example the energy density on flat slices

\[ \delta \rho_1 |_{\text{flat}} = \delta \rho_1 + \rho_0' \frac{\psi_1}{\mathcal{H}}. \]  
(3.31)

This is gauge invariant in the \( \xi^\mu \)-independence sense, but it does depend on the choice of background (e.g. a background depending on \( x^i \) instead of just time as in FRW would obviously give a very different result). This works for all the perturbations and also at second order and higher.

We conclude the above example by observing that to remove the gauge modes on subhorizon scales, often referred to as specifying the threading, we can choose \( \tilde{E}_1 = 0 \), which gives

\[ \beta_1 = -E_1. \]  
(3.32)

For the vector modes we choose \( \tilde{F}^i_1 = 0 \), which gives

\[ \gamma^i_1 = -F^i_1. \]  
(3.33)

Hence in this gauge the spatial part of the perturbed metric is zero with the exemption of the tensor modes.

3.2.3. Second order. At second order the generating vector \( \xi^\mu_2 \) is split into a scalar time and scalar and vector spatial part, similarly as at first order,

\[ \xi^\mu_2 = (\alpha_2, \beta_2, \gamma^i_2), \]  
(3.34)

where the vector part is divergence free \( \partial_k \gamma^k_2 = 0 \). We then find from equations (2.17) that a 4-scalar transforms at second order

\[ \tilde{\delta} \rho_2 = \delta \rho_2 + \rho_0' \alpha_2 + \alpha_1 (\rho_0'' \alpha_1 + \rho_0' \alpha'_1 + 2 \delta \rho'_1) + (2 \delta \rho_1 + \rho_0' \alpha_1) \tilde{\chi} (\beta_1, \gamma^i_2). \]  
(3.35)
We see here already the coupling between vector and scalar perturbations in the last term through the gradient and $\gamma_1'$. The gauge is only specified once the scalar temporal gauge perturbations, $\beta_1$ and $\gamma_1'$, are specified.

The metric tensor transforms at second order, from equation (2.17) as

$$
\delta g^{(2)}_{\mu \nu} = \delta g^{(0)}_{\mu \nu} + \delta g^{(0)}_{\mu \nu, i} \xi^i + \delta g^{(0)}_{\mu \nu, \lambda} \xi^\lambda + \delta g^{(1)}_{\mu \nu, \lambda, k} \xi^k + \delta g^{(1)}_{\mu \nu, \lambda, i} \xi^i + \delta g^{(1)}_{\mu \nu, \lambda, i, j} \xi^i \xi^j + 2 \left[ \delta g^{(1)}_{\mu \nu, \lambda, k} \xi^k + \delta g^{(1)}_{\mu \nu, \lambda, l, i} \xi^i \xi^l \right] + \delta g^{(1)}_{\mu \nu, \lambda, k, l} \xi^k \xi^l + \delta g^{(1)}_{\mu \nu, \lambda, i, j, k} \xi^i \xi^j \xi^k \right]
$$

Now following similar lines as at first order in the previous section, we could get the transformation behaviour for the second-order lapse function $\phi_2$ straight from the $0 \rightarrow 0$-component of equation (3.36).

Instead, to keep the discussion as brief as possible, we now turn to the transformation behaviour of the perturbations in the spatial part of the metric tensor. Here we can follow a similar procedure as in the linear case. But, the task is made more complicated not only by the size of the expressions but more importantly by the fact that now we will have to let inverse Laplacians operate on products, in order to get the transformations of the scalar, vector and tensor parts of the spatial metric.

Using equation (3.36) we find that the perturbed spatial part of the metric, $C_{2ij}$, transforms at second order as

$$
2 \hat{C}_{2ij} = 2 C_{2ij} + 2 \hat{H} \alpha \delta_{ij} + \xi_{2ij} + \xi_{2,ij} + X_{ij},
$$

where we defined $X_{ij}$ to contain the terms quadratic in the first-order perturbations as

$$
X_{ij} = 2 \left[ \hat{H}^2 + \frac{\alpha''}{a} \right] \alpha_1^2 + \hat{H} \left( \alpha_1 \alpha_1' + \alpha_1 k, \xi_k \right) \right] \delta_{ij} + 2 \left[ \alpha_1 \left( C_{2ij} + 2 \hat{H} C_{1ij} \right) 
$$

$$
+ C_{1ij,k} \xi^k + C_{1jk} \xi^i \xi^k + 2 \left( B_{1i} \alpha_1 + B_{1j} \alpha_1, i + B_{1j} \alpha_1, ji + 4 \hat{H} \alpha_1 \xi_{ii,j} + \xi_{ij,i} - 2 \alpha_1, i \alpha_1, j + 2 \xi_{i} \xi_{j} + \alpha_1 \left( \xi_{ii,j} + \xi_{ij,i} + \xi_{ij,i} + \xi_{ij,i} \right) \xi_j \right)
$$

$$
+ \xi_{ii,j} \xi_{i},j + \xi_{ij,i} \xi_{j},i + \xi_{ij,i} \alpha_1, j + \xi_{ij,i} \alpha_1, j, \alpha_1, j.
$$

Note that in equation (3.38) and in the following we will not decompose the spatial part of equation (3.11), $\xi_1 = \beta_1 + \gamma_1', \gamma_1'$, whenever convenient to keep the presentation as compact as possible.

The perturbed spatial part of the metric, $C_{2ij}$, is decomposed in equation (3.6) into scalar, vector and tensor parts, which we reproduce here at second order,

$$
2 C_{2ij} = - 2 \tilde{\psi}_2 \delta_{ij} + 2 \tilde{E}_{2ij} + 2 \tilde{F}_{2(ij)} + h_{2ij}.
$$

Taking the trace of equation (3.37) and substituting into equation (3.39) we get

$$
- \tilde{\psi}_2 + \nabla^2 \tilde{E}_2 = - \tilde{\psi}_2 + \nabla^2 \tilde{E}_2 + 3 \hat{H} \alpha_2 + \nabla^2 \beta_2 + \frac{1}{2} \lambda^2 \xi_2,
$$

where we find $\lambda_{ij}$ to be

$$
\frac{1}{2} \lambda_{ij} = 3 \left[ \hat{H}^2 + \frac{\alpha''}{a} \right] \alpha_1^2 + 3 \hat{H} \left( \alpha_1 \alpha_1' + \alpha_1 k, \xi_k \right) \right] \delta_{ij} + 2 \left[ \alpha_1 \left( C_{1ij} + 2 \hat{H} C_{1ij} \right) 
$$

$$
+ C_{1ik,j} \xi^k + C_{1jk} \xi^i \xi^k + 2 \left( B_{1i} \alpha_1 + B_{1j} \alpha_1, i + B_{1j} \alpha_1, ji + 4 \hat{H} \alpha_1 \xi_{ii,j} + \xi_{ij,i} - 2 \alpha_1, i \alpha_1, j + 2 \xi_{i} \xi_{j} + \alpha_1 \left( \xi_{ii,j} + \xi_{ij,i} + \xi_{ij,i} + \xi_{ij,i} \right) \xi_j \right)
$$

$$
+ \xi_{ii,j} \xi_{i},j + \xi_{ij,i} \xi_{j},i + \xi_{ij,i} \alpha_1, j + \xi_{ij,i} \alpha_1, j, \alpha_1, j.
$$

Now applying the operator $\delta' \delta'$ to equation (3.37) we get a second equation relating the scalar perturbations $\psi_2$ and $E_2$,

$$
- \tilde{\psi}_2 + \nabla^2 \tilde{E}_2 = - \tilde{\psi}_2 + \nabla^2 \tilde{E}_2 + 3 \hat{H} \alpha_2 + \nabla^2 \beta_2 + \frac{1}{2} \lambda_{ij}^2
$$
This gives for the transformations of the curvature perturbation at second order,
\[ \tilde{\psi}_2 = \psi_2 - \mathcal{H} \alpha_2 - \frac{1}{4} \lambda^h_k + \frac{1}{2} \nabla^2 \chi^{ij,ij}, \tag{3.43} \]
and for the shear scalar,
\[ \tilde{E}_2 = E_2 + \beta_2 + \frac{3}{2} \nabla^2 \chi^{ij,ij} - \frac{1}{4} \nabla^2 \chi^h_k. \tag{3.44} \]
Taking the divergence of equation (3.37) we get
\[ 2 \tilde{C}_{2ij,}^l = 2 C_{2ij,}^l + 2 \mathcal{H} \alpha_{2,ij} + \nabla^2 \xi_2 + \nabla^2 \beta_{2,ij} + \chi_{ik}^h. \tag{3.45} \]
Substituting in our results for \( \tilde{\psi}_2 \) and \( \tilde{E}_2 \) we then arrive at
\[ \nabla^2 \tilde{F}_{2i} = \nabla^2 F_{2i} + \nabla^2 \gamma_{2i} + \chi_{ik}^h - \nabla^2 \chi^{kl,kl}. \tag{3.46} \]
Finally
\[ \tilde{F}_{2i} = F_{2i} + \gamma_{2i} + \nabla^2 \chi_{ik}^h - \nabla^2 \chi^{kl,kl}. \tag{3.47} \]
We finally turn to the tensor perturbation at second order. Substituting our previous results for \( \tilde{\psi}_2 \), \( E_2 \) and \( F_2 \) into equation (3.37) we get, probably surprisingly,
\[ \tilde{h}_{2ij} = h_{2ij} + \chi_{ij} + \frac{1}{2} \left( \nabla^2 \chi^{kl,kl} - \chi^h_k \right) \delta_{ij} + \frac{1}{2} \nabla^2 \chi^{kl,kl}_{ij} \]
\[ + \frac{1}{2} \nabla^2 \chi^h_{k,ij} - \nabla^2 \left( \chi^{h,k}_{ij} + \chi^h_{jk} \right). \tag{3.48} \]
Although the second-order tensor transformation \( h_{2ij} \) is not dependent on the second-order gauge functions \( \xi^k_{ij} \), it does depend on first-order quantities quadratically.

The same holds for other quantities that are zero in the background: the first-order quantity is gauge invariant by virtue of the Stewart–Walker lemma [2] (and by construction). However the second-order quantity is no longer gauge invariant, as shown above in the case of the tensor perturbation, \( h_{2ij} \). This is not a violation of the Stewart–Walker lemma, it merely shows that the second-order quantities "live" in a first-order "background". Another example is the anisotropic stress, which is gauge invariant at first order, but not at second.

### 3.2.4. Constructing gauge-invariant variables at second order.

We can now construct, just as at first order in section 3.2.2, gauge-invariant variables at second order. We choose the same example as in the previous section, namely the energy density on flat hypersurfaces. But now also give the second-order tensor perturbation in this gauge.

To specify the gauge at second order we choose hypersurfaces on which \( \tilde{\psi}_2 = 0 \), and we see from equation (3.43) that gives
\[ \alpha_{2,\text{flat}} = \frac{\psi_2}{\mathcal{H}} + \frac{1}{4 \mathcal{H}} \left[ \nabla^2 \chi^{ij,ij}_{\text{flat},ij} - \chi^h_{\text{flat},k} \right]. \tag{3.49} \]
where we get \( \chi^h_{\text{flat},ij} \) from equation (3.38) using the first-order gauge generators given above as
\[ \chi^h_{\text{flat},ij} = 2 \left[ \psi_1 \left( \frac{\psi_1'}{\mathcal{H}} + 2 \psi_1 \right) \psi_{1,k} \xi^k_{\text{flat}} \right] \delta_{ij} + \frac{4}{\mathcal{H}} \psi_1 \left( C^h_{1,ij} + 2 \mathcal{H} C_{1,ij} \right) \]
\[ + 4 C_{1,ij,k} \xi^k_{\text{flat}} + \left( 4 C_{1,ik} + \xi^i_{\text{flat},k} \right) \xi^k_{\text{flat},l} + \left( 4 C_{1,jk} + \xi^j_{\text{flat},k} \right) \xi^k_{\text{flat},l} \]
\[ + \frac{1}{\mathcal{H}} \left[ \psi_1 \left( 2 B_{1,j} + \xi^j_{\text{flat},l} \right) + \psi_1 \left( 2 B_{1,i} + \xi^i_{\text{flat},l} \right) \right] - \frac{2}{\mathcal{H}^2} \psi_{1,j} \psi_{1,j} \]
\[ + \frac{2}{\mathcal{H}} \psi_1 \left( \xi^i_{\text{flat},i} + 4 \mathcal{H} \xi^i_{\text{flat},j} \right) + 2 \xi^i_{\text{flat},i} \xi^i_{\text{flat},j} + 2 \xi^i_{\text{flat},i} \xi^i_{\text{flat},j}, \tag{3.50} \]
where we define
\[ \xi^i_{\text{flat},ij} = -(E_{1,i} + F_{1,i}). \tag{3.51} \]
The trace of equation (3.50) is then
\[ X_{\text{flat}}^k = 6 \left[ \psi_1 \left( \frac{\psi_1}{\mathcal{H}} + 2 \psi_1 \right) + \psi_{1,k} \xi_1^{\text{flat}} \right] + \frac{4 \mathcal{H}}{\psi_1} \left( C_{1,k}^k + 2 \mathcal{H} C_{1,k}^k \right) + 4 C_{1,k}^k \xi_1^{\text{flat}} + 4 (2 C_{1,k}^k + \xi_1^{\text{flat}}) \xi_1^{\text{flat},(k,i)} - 2 \mathcal{V}^2 E_{1,k} \xi_1^{\text{flat}} + \frac{2}{\mathcal{H}} \left( 2 B_{1k} + \xi_1^{\text{flat},k} - \frac{1}{\mathcal{H}} \psi_{1,k} \right) \psi_1,^k - \frac{2}{\mathcal{H}} (\psi_1 \mathcal{V}^2 E_1 + 4 \mathcal{H} \mathcal{V}^2 E_1). \] (3.52)

Then substituting at first-order equations (3.30), (3.32) and (3.33), and at second-order equation (3.49) into equation (3.35), we get the second-order energy density perturbation on uniform curvature hypersurfaces [10]
\[ \tilde{\delta} \rho_{2,\text{flat}} = \delta \rho_2 + \frac{\rho_0'}{\mathcal{H}} \psi_2 + \frac{\rho_0'}{4 \mathcal{H}} \left( \mathcal{V}^2 \chi_{\text{flat},ij}^{jj} - \chi_{\text{flat},k}^k \right) + \frac{\psi_1}{\mathcal{H}^2} \left[ \rho_0' \psi_1 + \rho_0' \left( \psi_1 - \frac{\mathcal{H}'}{\mathcal{H}} \psi_1 \right) + 2 \mathcal{H} \delta \rho_1 \right] + \left( 2 \mathcal{H} \delta \rho_1 + \frac{\rho_0'}{\mathcal{H}} \psi_1 \right) \xi_1^{\text{flat}}. \] (3.53)

The second-order tensor perturbation in that the flat gauge, i.e. on uniform curvature hypersurfaces, is given by substituting equations (3.30), (3.32) and (3.33) into equation (3.48), and we find after some algebra
\[ \tilde{h}_{2,\text{flat},ij} = h_{2,ij} + \chi_{\text{flat},ij}^k + \frac{1}{2} \left( \mathcal{V}^2 \chi_{\text{flat},kl}^{kl} - \chi_{\text{flat},k}^k \right) \delta_{ij} + \frac{1}{2} \mathcal{V}^2 \mathcal{V}^2 \chi_{\text{flat},kl}^{kl} - \mathcal{V}^2 \left( \chi_{\text{flat},k}^k + \chi_{\text{flat},j}^k \right). \] (3.54)

4. Discussion and conclusions

This is neither the first nor will it be the last discussion of perturbation theory in cosmology. However, in this concise introduction we have tried to strike a balance between mathematical rigour and ease of application of the results. For a more detailed exposition of cosmological perturbation theory and further references see [21].

We have here studied perturbations about a flat FRW background spacetime. But as pointed out above, the formalism introduced by Bardeen can easily be applied to other settings and background spacetimes, and can also be extended beyond GR. Indeed, perturbation theory and the formalism discussed in this paper can be applied to all covariant metric theories. Although here we have assumed standard four-dimensional (4D) Einstein gravity throughout, the formalism has also been applied, for example, to 5D braneworld models (see, e.g., [22] for an overview), and has been used to construct gauge-invariant variables in that theory.

Whereas most of the material discussed in the previous sections has been expounded elsewhere, albeit often in different form and with other aims, we are not aware of the derivation of how the second-order tensor perturbations transforms in full generality under gauge transformations being discussed elsewhere (see however [23] for the case of scalar perturbations). Also its representation in the uniform curvature gauge has been discussed for the first time. These results will be of particular interest in second-order calculations of the gravitational wave background [24]. The transformations at second order of the decomposed components of the spatial parts of the metric have also not been discussed in the literature before, and will be particularly useful in relating quantities calculated in different gauges.

Acknowledgments

The authors are grateful to Adam Christopherson, David Seery and David Wands for useful discussions and comments.
Appendix. Definitions

In this appendix, we bring together some key ideas and definitions from differential geometry, in particular those regarding maps of manifolds, which are useful in setting up perturbation theory in general relativity where covariance matters. This appendix is not a comprehensive study as our intention is only to show, intuitively, why certain expressions take the form they do and to define some key ideas. In order to be useful to those who do not wish to enter into the formalism of differential geometry we will use coordinate expressions where possible. For those who worry about these things we will therefore be working in coordinate neighbourhoods and all functions will be assumed to be adequately differentiable as they are for most of cosmology. For more details and a less coordinate dependent approach the reader is referred to the books by Hawking and Ellis [25] and by Wald [26].

A.1. Maps between manifolds

We assume some familiarity with the definition and properties of differential manifolds and are concerned here with maps between such manifolds, in particular a diffeomorphism. Consider two manifolds $\mathcal{M}$ and $\mathcal{M}_e$ and denote the chart maps (coordinates) on each by

$$f_a : O_a \rightarrow U_a \quad \text{for} \quad O_a \subset \mathcal{M}, U_a \subset \mathbb{R}^n,$$

$$g_b : O_b \rightarrow U_b \quad \text{for} \quad O_b \subset \mathcal{M}_e, U_b \subset \mathbb{R}^n.$$  \hspace{1cm} (A.1)

In words the function $f_a$ assigns coordinates (in $U_a$) to points in the $n$-dimensional neighbourhood $O_a$ of $\mathcal{M}$ and $g_b$ does the same in the neighbourhood $O_b$ of $\mathcal{M}_e$.

The map

$$\phi : \mathcal{M} \rightarrow \mathcal{M}_e \quad \hspace{1cm} \text{(A.2)}$$

is $C^\infty$, i.e. infinitely continuously differentiable in the advanced calculus sense, if for each $a$ and $b$ the map

$$f_a \circ \phi^{-1} \circ g_b^{-1} : U_b \rightarrow U_a \quad \hspace{1cm} \text{(A.3)}$$

is $C^\infty$. The map $\phi$ is a diffeomorphism if it is one to one, onto and $\phi$ and its inverse $\phi^{-1}$ is $C^\infty$. Loosely speaking, for coordinates $x^i$ on $U_a$ and $y^i$ on $U_b$, equation (A.3) relating the coordinates under the map $\phi$ can be represented by the equations

$$y^\mu = \phi^\mu(x^\nu), \quad \hspace{1cm} \text{(A.4)}$$

with inverse

$$x^\mu = (\phi^{-1})^\mu(y^\nu), \quad \hspace{1cm} \text{(A.5)}$$

where $\phi^\mu$ and $(\phi^{-1})^\mu$ are $C^\infty$.

A.2. Maps of vectors induced by mapping manifolds

A manifold map $\phi : \mathcal{M} \rightarrow \mathcal{M}_e$ induces a map of the tangent vectors at $p$ on $\mathcal{M}$ to tangent vectors at $\phi(p)$ on $\mathcal{M}_e$. We will write this map as $\phi^* : V_p \rightarrow V_{\phi(p)}$ where $V_p$ and $V_{\phi(p)}$ denote the tangent spaces at $p$ and $\phi(p)$. Using the coordinate description above, the map $\phi^*$ can be written as

$$(\phi^*)^\nu_\mu = \frac{\partial \phi^\nu}{\partial x^\mu}, \quad \hspace{1cm} \text{(A.6)}$$

and if $X^\mu \in V_p$ and $Y^\nu \in V_{\phi(p)}$ then

$$(\phi^*)^\nu_\mu : X^\nu \rightarrow Y^\mu, \quad \text{or} \quad Y^\mu = (\phi^*)^\mu_\nu X^\nu = \frac{\partial \phi^\mu}{\partial x^\nu} X^\nu, \quad \hspace{1cm} \text{(A.7)}$$

in the usual notation. The map $\phi^*$ is sometimes called a pushforward.
A.3. Pullback

The concept of a pullback is well described by its name. If we are given a function \( f \) defined on a manifold \( N \) so
\[
f : N \to \mathbb{R}^n,
\]
and a manifold map
\[
\phi : M \to N,
\]
then \( \phi \) can be used to pullback \( f \) from \( N \) to \( M \) by using the composite map \( f \circ \phi \). This pulls \( f \) back from \( N \) to map \( M \) to \( \mathbb{R}^n \), i.e.,
\[
f \circ \phi : M \to \mathbb{R}^n.
\]
This is perhaps easier to follow if viewed as a sequence: the \( \phi \) maps \( M \) to \( N \) then \( f \) maps \( N \) to \( \mathbb{R}^n \) or, in symbols,
\[
f \circ \phi : M \longrightarrow f : N \longrightarrow \mathbb{R}^n.
\]
The map \( f \circ \phi \) is the pullback of \( f \).

A.4. Maps of covectors induced by mapping manifolds

For covectors (forms or covariant vectors) the map between manifolds \( \phi : M \to N \) induces a pullback map \( \phi^* \) which takes covectors at \( \phi(p) \) on \( N \) to covectors at \( p \) on \( M \), i.e.,
\[
\phi^* : W_{\phi(p)} \to W_p,
\]
where \( W_{\phi(p)} \) denotes the cotangent space at \( \phi(p) \) on \( N \) and \( W_p \) is the cotangent space at \( p \) on \( M \). This can be written in coordinates, by letting \( X_\mu \in W_p \) and \( Y_\nu \in W_{\phi(p)} \), and we can write
\[
X_\mu = (\phi^*)_\mu^\nu Y_\nu = \frac{\partial \phi^\nu}{\partial x_\mu} Y_\nu, \quad \text{or} \quad Y_\nu = \left( \frac{\partial \phi^\nu}{\partial x_\mu} \right)^{-1} X_\mu.
\]
The mapping of covariant and contravariant tensors of higher rank follows the same pattern as that for the vectors and covectors. Tensors with mixed co- and contravariant indices also follow this pattern, although proving it requires a little care (see [26, p 438]). A pullback on a vector field effectively reverses the effect of a pushforward and so the definitions of pullback and pushforward act to preserve the scalar property of the product of a vector and a covector.

A.5. Pullback of a composite map

Note that for composites of maps the pullback behaves in a different way to the push forward. For pushforward maps,
\[
\phi : M \to N, \quad \text{and} \quad \psi : N \to P,
\]
the composite is simply
\[
\psi \circ \phi : M \to P.
\]
For the corresponding pullback maps we have
\[
\phi_* : M \to \mathcal{M}, \quad \psi_* : \mathcal{P} \to \mathcal{N},
\]
\[
(\psi \circ \phi)_* : \mathcal{P} \to \mathcal{M},
\]
but, and this is important, if we remove the bracket we have to reverse the maps
\[
(\psi \circ \phi)_* = \phi_* \circ \psi_* \quad \text{as can be seen by looking at the actions of the maps themselves}.
\]
A.6. Lie derivative

Let $\mathcal{M}$ be a differential manifold and let $\xi$ be a vector field on $\mathcal{M}$, then $\xi$ generates a map of $\mathcal{M}$ onto itself as follows. In a coordinate neighbourhood, solve the system of ordinary differential equations

$$\frac{dx^\mu}{d\epsilon} = \xi^\mu$$

in $\mathbb{R}^n$. Given initial points in $\mathcal{M}$ at $\epsilon = 0$ this will generate a unique one-parameter family of integral curves with one and only one curve through each point in a neighbourhood. For an integral curve $\gamma$ let $\phi_\epsilon(p)$ be the point a distance $\epsilon$ along $\gamma$ from $p$. Applied to the family of curves $\phi_\epsilon$ generates a one-parameter family of diffeomorphisms of $\mathcal{M}$ onto itself. Clearly $\phi_{\epsilon_1} \circ \phi_{\epsilon_2} = \phi_{\epsilon_1 + \epsilon_2}$ and $\phi_0$ is the identity map.

Now let $T$ be a tensor field on $\mathcal{M}$ then the pullback $\phi_\epsilon^*$ of $\phi_\epsilon$ defines a new tensor field $\phi_\epsilon^*T$ on $\mathcal{M}$ which is a function of $\epsilon$. This enables us to define the Lie derivative, a covariant differentiation on $\mathcal{M}$, which does not increase the rank of the tensor,

$$\mathcal{L}_\xi T := \lim_{\epsilon \to 0} \frac{1}{\epsilon} (\phi_\epsilon^* T - T).$$

(A.20)

This can be interpreted as follows—the map $\phi_\epsilon^*$ pulls back the value of $T$ at $\phi_\epsilon(p)$ to $p$ from which we subtract the actual value of $T$ at $p$. This difference is a well-defined tensor quantity since the difference is taken at the point $p$. We can now take the limit to obtain a meaningful derivative at $p$—called the Lie derivative.

A.7. Exponential operator

Given an operator $A$ the exponential operator is defined by the formal series

$$e^A := 1 + A + \frac{1}{2} A^2 + \sum_{n=3}^{\infty} \left( \frac{1}{n!} \right) A^n.$$

(A.21)

Thus, for instance, if $A = \epsilon \mathcal{L}_\xi$ is the Lie derivative operator multiplied by $\epsilon$ the corresponding exponential operator is given by

$$e^{\epsilon \mathcal{L}_\xi} := 1 + \epsilon \mathcal{L}_\xi + \frac{\epsilon^2}{2} \mathcal{L}_\xi^2 + \sum_{n=3}^{\infty} \left( \frac{\epsilon^n}{n!} \right) \mathcal{L}_\xi^n.$$

(A.22)

References

[1] Sachs R K 1964 Relativity, Groups and Topology ed C De Witt and B De Witt (New York: Gordon and Breach)
[2] Stewart J M and Walker M 1974 Proc. R. Soc. Lond. A 341 49
[3] Stewart J M 1990 Class. Quantum Grav. 7 1169
[4] Mukhanov V F, Feldman H A and Brandenberger R H 1992 Phys. Rep. 215 203
[5] Bardeen J M 1980 Phys. Rev. D 22 1882
[6] Kodama H and Sasaki M 1984 Prog. Theor. Phys. Suppl. 78 1
[7] Durrer R 1994 Fund. Cosmic Phys. 15 209 (Preprint astro-ph/9311041)
[8] Bruni M, Matarrese S, Mollerach S and Sonego S 1997 Class. Quantum Grav. 14 2585 (Preprint gr-qc/9609040)
[9] Mukhanov V F, Abramo L R W and Brandenberger R H 1997 Phys. Rev. Lett. 78 1624 (Preprint gr-qc/9609026)
[10] Malik K A and Wands D 2004 Class. Quantum Grav. 21 L65 (Preprint astro-ph/0307055)
[11] Noh H and Hwang J C 2004 Phys. Rev. D 69 104011
[12] Bartolo N, Komatsu E, Matarrese S and Riotto A 2004 Phys. Rev. D 70 103 (Preprint astro-ph/0406398)
[13] Nakamura K 2007 Prog. Theor. Phys. 117 17 (Preprint gr-qc/0605108)
[14] Ellis G F R and Bruni M 1989 Phys. Rev. D 40 1804
[15] Langlois D and Vernizzi F 2005 *Phys. Rev.* D **72** 103501 (Preprint astro-ph/0509078)
[16] Sopuerta C F, Bruni M and Gualtieri L 2004 *Phys. Rev.* D **70** 064002 (Preprint gr-qc/0306027)
[17] Salopek D S and Bond J R 1990 *Phys. Rev.* D **42** 3936
[18] Lyth D H, Malik K A and Sasaki M 2005 *J. Cosmol. Astropart. Phys.* JCAP05(2005)004 (Preprint astro-ph/0411220)
[19] Malik K A 2001 Preprint astro-ph/0101563
[20] Malik K A and Wands D 2005 *J. Cosmol. Astropart. Phys.* JCAP02(2005)007 (Preprint astro-ph/0411703)
[21] Malik K A and Wands D 2008 in preparation
[22] Maartens R 2004 *Living Rev. Rel.* 7 7 (Preprint gr-qc/0312059)
[23] Pitran C 2007 *Class. Quantum Grav.* **24** 6127 (Preprint arXiv:0706.4383)
[24] Ananda K N, Clarkson C and Wands D 2007 *Phys. Rev.* D **75** 123518 (Preprint gr-qc/0612013)
[25] Hawking S W and Ellis G F R 1973 *The Large Scale Structure on Space-Time* (Cambridge: Cambridge University Press)
[26] Wald R M 1984 *General Relativity* (Chicago, IL: University of Chicago Press)