A New Framework for Treating Small Scale Inhomogeneities in Cosmology

Robert M. Wald

with Stephen Green

arXiv:1011.4920; Phys.Rev. **D83**, 084020 (2011)

Usual Assumptions in Cosmology

It is generally believed that our universe is very well described on large scales by a Friedmann-Lemaitre-Robertson-Walker (FLRW) model. The FLRW models treat the matter as being homogeneously distributed. However, on small scales, extremely large departures of the mass density from FLRW models are commonly observed, e.g., for the Earth $\delta\rho/\rho \sim 10^{30}$. Thus, at least with regard to the description of matter, the FLRW models would seem to provide a very poor description of our universe on small scales.

Nevertheless, common sense estimates suggest that (a) the deviation of the metric (as opposed to mass density, which corresponds to second derivatives of the metric) from an FLRW metric are globally very small on all scales except in the immediate vicinity of strong field objects such as black holes and neutron stars, and (b) the terms in Einstein's equation that are nonlinear in the deviation of the metric from a FLRW metric are negligibly small as compared with the dominant linear terms in the deviation from a FLRW metric except in the immediate vicinity of strong field objects.

These common sense estimates together with the fact that the motion of matter relative to the rest frame of the cosmic microwave background is non-relativistic strongly suggest that (1) the large scale structure of the universe is well described by a FLRW metric, (2) when averaged on scales sufficiently large that $|\delta\rho/\rho| \ll 1$—i.e., scales $\gg 10$ Mpc in the present universe—the deviations from a FLRW model are well described by ordinary FLRW linear perturbation theory, and (3) on smaller scales, the deviations from a FLRW model (or, for that matter, from Minkowski spacetime) are well described by Newtonian gravity—except, of course, in the immediate vicinity of strong field objects.

These assumptions underlie the standard cosmological model, which has been remarkably successful in accounting for essentially all cosmological phenomena. Thus, there is good empirical evidence that assumptions (1)–(3) are at least essentially correct.

Unsatisfactory Aspects of Cosmological Assumptions

If correct, one would like to derive the above properties in a systematic and mathematically satisfactory way, not merely by making "common sense estimates". Indeed, it is not even obvious that assumptions (1)–(3) are mathematically consistent: It is clear that one would get an extremely poor description of small-scale structure in the universe if one neglected the nonlinear terms in Einstein's equation in the deviation of the metric from a FLRW model; for example, galaxies would not be bound. But if one cannot neglect nonlinear terms in Einstein's equation on small scales, how can one justify neglecting them on large (i.e., $\gg 10$ Mpc) scales?

Indeed, it is far from obvious, *a priori*, that nonlinearities associated with small-scale inhomogeneities could not produce important effects on the large-scale dynamics of the FLRW model itself, as has been suggested by a number of authors as a possible way to account for the effects of "dark energy" without invoking a cosmological constant, a new source of matter, or a modification of Einstein's equation.

In addition, since it is not clear exactly what approximations are needed for assumptions (1)–(3) to be valid, it is far from clear as to how one could go about systematically improving these approximations.

Averaging Over Inhomogeneities

The main approach that has been taken thus far (by Buchert and others) is to consider inhomogeneous models, take spatial averages to define corresponding FLRW quantities, and derive equations of motion for these FLRW quantities. Since, in particular, the spatial average of the square of a quantity does not equal the square of its spatial average, the effective FLRW dynamics of an inhomogeneous universe will differ from that of a homogeneous universe. However, there are a number of serious difficulties with this approach:

- It is not obvious how to interpret the averaged quantities in terms of observable quantities. For example, if the total volume of a spatial region is found to increase with time, this certainly does not imply that observers in this region will find that Hubble's law appears to be satisfied.

- The notion of averaging is slicing dependent and the average of tensor quantities over a region in a non-flat spacetime is intrinsically ill defined.
- The equations for averaged quantities that have been derived to date are only a partial set of equations—they contain quantities whose evolution is not determined—so it is difficult to analyze what dynamical behavior of the averaged quantities is actually possible.

The Type of Framework We Seek

We seek a framework that allows spacetimes where there can be significant inhomogeneity and nonlinear dynamics on small scales, but can describe "average" large-scale behavior in a mathematically precise manner, with approximations that are "controlled" in the sense that they hold with arbitrarily good accuracy in some appropriate limit. The key elements in this framework are:

- There is a "background spacetime metric", $g^{(0)}_{ab}$, that is supposed to correspond to the metric "averaged" over small scale inhomogeneities. The difference, $h_{ab} \equiv g_{ab} - g^{(0)}_{ab}$, between the actual metric g_{ab} and the background metric is assumed to be "small".

- Although h_{ab} is "small", spacetime derivatives of h_{ab} are *not* assumed to be small. In particular, quadratic products of $\nabla_c h_{ab}$ are allowed to be of the same order as the curvature of $g^{(0)}_{ab}$. This allows nonlinear terms in Einstein's equation to affect the dynamics of the background metric.

- No restrictions are placed upon second derivatives of h_{ab}. In particular, if matter is present, we allow $\delta\rho/\rho \gg 1$.

How to Make Our Framework Precise

In order to develop a mathematically precise framework, we wish to consider a one-parameter family of metrics $g_{ab}(\lambda)$ that has appropriate limiting behavior as $\lambda \to 0$. In our case, we want the "small parameter" λ to be related to the ratio of the lengthscale associated with the small-scale inhomogeneities to the curvature lengthscale of the "background metric" $g_{ab}^{(0)} \equiv g_{ab}(\lambda = 0)$.

Example: Ordinary Perturbation Theory: Here the "small parameter" λ is simply the amplitude of the deviation, h_{ab}, of the metric from the background metric. Spacetime derivatives of h_{ab} are assumed to be correspondingly small. To implement this in a mathematically precise way, we consider a one-parameter family of metrics $g_{ab}(\lambda, x)$ that is jointly smooth in the parameter λ and the spacetime coordinates x. If $g_{ab}(\lambda)$ satisfies Einstein's equation for all $\lambda > 0$, then $g_{ab}^{(0)}$ also automatically satisfies Einstein's equation. Define the nth order perturbation $g_{ab}^{(n)} \equiv (\partial^n g_{ab}/\partial\lambda^n)|_{\lambda=0}$. It satisfies an equation obtained by taking the nth partial derivative with respect to λ at $\lambda = 0$ of Einstein's equation for $g_{ab}(\lambda, x)$.

Our Framework: As in ordinary perturbation theory we want to consider a one-parameter family $g_{ab}(\lambda)$ that approaches a "background metric" $g_{ab}^{(0)}$ as $\lambda \to 0$. However, we do *not* want spacetime derivatives of $g_{ab}(\lambda)$

to approach corresponding derivatives of $g_{ab}^{(0)}$ as $\lambda \to 0$.

Can this be made mathematically consistent?

Yes! The issues we face are very similar to the issues arising when one attempts to treat the self-gravitating effects of short-wavelength gravitational radiation. We will adopt a version of Burnett's formulation of the "shortwave approximation," which we generalize to allow for the presence of a nonvanishing matter stress-energy tensor T_{ab}.

Weak Limits

We want to consider a limit in which $h_{ab} \equiv g_{ab}(\lambda) - g^{(0)}_{ab}$ becomes small as $\lambda \to 0$, but $\nabla_c h_{ab}$ does not become small. A prototype example of the kind of behavior we want to allow is

$$h(x) = \lambda \sin(x/\lambda)$$

Then $h \to 0$ as $\lambda \to 0$ but $\nabla h \sim \cos(x/\lambda)$ does not approach a limit in the ordinary (uniform or pointwise) sense. However, it does approach a limit in the *weak* sense:

Definition: Let $A_{a_1 \dots a_n}(\lambda)$ be a one-parameter family of tensor fields defined for $\lambda > 0$. We say that $A_{a_1 \dots a_n}(\lambda)$

converges weakly to $B_{a_1\dots a_n}$ as $\lambda \to 0$ if for all smooth $f^{a_1\dots a_n}$ of compact support, we have

$$\lim_{\lambda\to 0}\int f^{a_1\dots a_n} A_{a_1\dots a_n}(\lambda) = \int f^{a_1\dots a_n} B_{a_1\dots a_n} \; .$$

Roughly speaking, the weak limit performs a local spacetime average of $A_{a_1\dots a_n}(\lambda)$ before letting $\lambda \to 0$.

In our above example, it is easy to see that $\cos(x/\lambda)$ converges weakly to zero. However, note that $(\nabla h)^2 \sim \cos^2(x/\lambda)$ converges weakly to $1/2$, not to 0. As we shall see, terms involving quadratic products of $\nabla_c h_{ab}$ will act as an "effective gravitational stress-energy tensor."

Our Assumptions

Let ∇_a denote an arbitrary fixed (i.e., λ-independent) derivative operator on the spacetime manifold M. For convenience in stating these conditions, we choose an arbitrary Riemannian metric e_{ab} on M and for any tensor field $t_{a_1 \ldots a_n}$ on M we define
$|t_{a_1 \ldots a_n}|^2 = e^{a_1 b_1} \ldots e^{a_n b_n} t_{a_1 \ldots a_n} t_{b_1 \ldots b_n}$.

(i) For all $\lambda > 0$, we have

$$G_{ab}(g(\lambda)) + \Lambda g_{ab}(\lambda) = 8\pi T_{ab}(\lambda) \ ,$$

where $T_{ab}(\lambda)$ satisfies the weak energy condition, i.e., for all $\lambda > 0$ we have $T_{ab}(\lambda) t^a(\lambda) t^b(\lambda) \geq 0$ for all vectors $t^a(\lambda)$ that are timelike with respect to $g_{ab}(\lambda)$.

(ii) There exists a smooth positive function $C_1(x)$ on M such that

$$|h_{ab}(\lambda, x)| \leq \lambda C_1(x) \ ,$$

where $h_{ab}(\lambda, x) \equiv g_{ab}(\lambda, x) - g_{ab}(0, x)$.

(iii) There exists a smooth positive function $C_2(x)$ on M such that

$$|\nabla_m h_{ab}(\lambda, x)| \leq C_2(x) \ .$$

(iv) There exists a smooth tensor field μ_{mnabcd} on M such that

$$\mathrm{wlim}_{\lambda \to 0} \left[\nabla_m h_{ab}(\lambda) \nabla_n h_{cd}(\lambda)\right] = \mu_{mnabcd} \ ,$$

where "wlim" denotes the weak limit.

It follows immediately that $\mu_{mn(ab)(cd)} = \mu_{mnabcd}$ and $\mu_{mnabcd} = \mu_{nmcdab}$, and it is not difficult to show that $\mu_{(mn)abcd} = \mu_{mnabcd}$. It also is not difficult to see that if $g_{ab}(\lambda)$ satisfies the above conditions for any choice of ∇_a and e_{ab}, then it satisfies these conditions for all choices of ∇_a and e_{ab}. In our calculations, it will be convenient to choose ∇_a to be the derivative operator associated with the background metric $g^{(0)}_{ab} \equiv g_{ab}(0)$, and in the following, we shall make this choice. We shall also raise and lower indices with $g^{(0)}_{ab}$.

Einstein's Equation

$$
\begin{aligned}
& R_{ab}(g^{(0)}) - \frac{1}{2} g_{ab}(\lambda) g^{cd}(\lambda) R_{cd}(g^{(0)}) + \Lambda g_{ab}(\lambda) \\
= \quad & 8\pi T_{ab}(\lambda) + 2\nabla_{[a} C^m{}_{m]b} - 2C^n{}_{b[a} C^m{}_{m]n} \\
& - g_{ab}(\lambda) g^{cd}(\lambda) \nabla_{[c} C^m{}_{m]d} + g_{ab}(\lambda) g^{cd}(\lambda) C^n{}_{d[c} C^m{}_{m]n}
\end{aligned}
$$

where

$$
C^c{}_{ab} = \frac{1}{2} g^{cd}(\lambda) \left\{ \nabla_a g_{bd}(\lambda) + \nabla_b g_{ad}(\lambda) - \nabla_d g_{ab}(\lambda) \right\}
$$

Take the weak limit as $\lambda \to 0$ of both sides of Einstein's equation. Get

$$
G_{ab}(g^{(0)}) + \Lambda g^{(0)}_{ab} = 8\pi T^{(0)}_{ab} + 8\pi t^{(0)}_{ab} \,,
$$

where $T^{(0)}_{ab} \equiv \mathrm{wlim}_{\lambda\to 0} T_{ab}(\lambda)$ (which necessarily exists) and

$$
\begin{aligned}
8\pi t^{(0)}_{ab} &= \frac{1}{8} g^{(0)}_{ab} \{ -\mu_c{}^c{}_{mn}{}^{mn} - \mu_c{}^c{}_m{}^m{}_n{}^n + 2\mu_{mn}{}^m{}_c{}^{cn} \} \\
&+ \frac{1}{2}\mu_{mn}{}^n{}_{ab}{}^m - \frac{1}{2}\mu_m{}^m{}_{nab}{}^n + \frac{1}{4}\mu_{abmn}{}^{mn} \\
&- \frac{1}{2}\mu_{m(ab)}{}^m{}_n{}^n + \frac{3}{4}\mu_m{}^m{}_n{}^n{}_{ab} - \frac{1}{2}\mu_{mn}{}^{mn}{}_{ab} \,.
\end{aligned}
$$

This expression is gauge invariant.

Tracelessness of $t^{(0)}_{ab}$

Multiply Einstein's equation by $h_{ef}(\lambda)$ and take the weak limit as $\lambda \to 0$. Obtain

$$\begin{aligned} \alpha_{amb}{}^{m}{}_{ef} &= 4\pi \,\mathrm{wlim}_{\lambda\to 0} h_{ef}(\lambda)[T_{ab}(\lambda) \\ &\quad -\frac{1}{2} g_{ab}(\lambda) g^{cd}(\lambda) T_{cd}(\lambda)] \,. \end{aligned}$$

where $\alpha_{abcdef} \equiv \mu_{[c|[ab]|d]ef}$. The right side can be proven to vanish if $T_{ab}(\lambda)$ satisfies the weak energy condition. From this, we immediately obtain

$$t^{(0)a}{}_{a} = 0 \,.$$

Positivity of Effective Gravitational Energy Density

Let t^a be timelike with respect to $t^{(0)}_{ab}$. We have

$$8\pi t^{(0)}_{ab} t^a t^b = \frac{1}{4}\left\{\mu_i{}^i{}_{jk}{}^{jk} - 2\mu_{jik}{}^{ijk} + 2\mu_{jki}{}^{ijk} - \mu_i{}^i{}_j{}^j{}_k{}^k\right\}.$$

where only spatial indices (orthogonal to t^a) appear on the right side. Let $P \in M$, choose Riemannian normal coordinates x about P, and let

$$\psi_{ab}(\delta, \lambda) \equiv f^\delta_P h_{ab}(\lambda).$$

where $f^\delta_P(x)$ is sharply peaked about P and its square approaches a δ-function as $\delta \to 0$. Then

$$\mu_{\mu\nu\alpha\beta\gamma\rho}(P) = \lim_{\delta\to 0}\lim_{\lambda\to 0}\int \partial_\mu\psi_{\alpha\beta}\partial_\nu\psi_{\gamma\rho} d^4x.$$

We obtain

$$t_{00}^{(0)}(P) = \frac{1}{32\pi} \lim_{\delta\to 0} \lim_{\lambda\to 0} \int d^4x [\partial_i \psi_{jk} \partial^i \psi^{jk} - 2\partial_j \psi_k{}^i \partial_i \psi^{jk} + 2\partial_j \psi_i{}^i \partial_k \psi^{jk} - \partial_i \psi_j{}^j \partial^i \psi_k{}^k] \,.$$

Now take the Fourier transform of ψ_{jk} and decompose it into its scalar, vector, and tensor parts

$$\hat{\psi}_{ij}(t, \boldsymbol{k}) = \hat{\sigma}(t, \boldsymbol{k}) k_i k_j + 2\hat{\phi} q_{ij} + 2k_{(i} \hat{z}_{j)}(t, \boldsymbol{k}) + \hat{s}_{ij}(t, \boldsymbol{k}) \,.$$

where $k^i \hat{z}_i = 0 = k^i \hat{s}_{ij}$, and $\hat{s}_i^i = 0$ and q_{ij} is the projection orthogonal to k^i of the Euclidean metric on Fourier transform space. The corresponding formula for

$t_{00}^{(0)}$ is

$$t_{00}^{(0)}(P) = \frac{1}{32\pi} \lim_{\delta\to 0} \lim_{\lambda\to 0} \int dt d^3\boldsymbol{k} \left\{ k_i k^i \hat{s}_{jk} \overline{\hat{s}^{jk}} - 8 k_i k^i \hat{\phi} \overline{\hat{\phi}} \right\} ,$$

Thus, the "tensor part," $\hat{s}^{ij}$, of $\hat{\psi}_{ij}$ ("gravitational radiation") contributes positive effective energy density, while the scalar part contributes negative energy density. From Einstein's equation, one can show that ϕ satisfies a Poisson-like equation. Its contribution to $t_{00}^{(0)}$ corresponds to (twice) the Newtonian formula for gravitational potential energy. By a fairly lengthy argument, this negative contribution to $t_{00}^{(0)}$ can be shown to vanish provided that $T_{ab}(\lambda)$ satisfies the weak energy condition.

Thus, $t_{ab}^{(0)}$ is traceless and satisfies the weak energy condition. It cannot provide any effects that mimic "dark energy."

Cosmological Perturbation Theory

The ordinary (uniform or pointwise) limit of $h_{ab}/\lambda = [g_{ab}(\lambda) - g^{(0)}_{ab}]/\lambda$ cannot exist for the one-parameter families of metrics of interest to us. However, its weak limit can exist, and, if it does, the resulting quantity

$$\gamma^{(L)}_{ab} \equiv \mathrm{wlim}_{\lambda\to 0}\frac{h_{ab}(\lambda)}{\lambda}$$

can be interpreted as the "long wavelength part" of the linear order in λ deviation of $g_{ab}(\lambda)$ from $g^{(0)}_{ab}$. We refer to

$$h^{(S)}_{ab}(\lambda) \equiv h_{ab}(\lambda) - \lambda\gamma^{(L)}_{ab}\,,$$

as the "short wavelength part" of the deviation of the metric from $g^{(0)}_{ab}$. If we divide Einstein's equation by λ, take the weak limit as $\lambda \to 0$, and if we assume that weak limits of various quantities such as

$$\mu^{(1)}_{abcdef} = \mathrm{wlim}_{\lambda\to 0}\frac{1}{\lambda}\left[\nabla_a h^{(S)}_{cd}(\lambda)\nabla_b h^{(S)}_{ef}(\lambda) - \mu_{abcdef}\right]$$

exist, then we obtain a linear equation for $\gamma^{(L)}_{ab}$ with a source term of the form

$$G^{(1)}_{ab}(g^{(0)}, \gamma^{(L)}) + \Lambda\gamma^{(L)}_{ab} + f_{ab}(g^{(0)}, \mu\gamma^{(L)}) = 8\pi T^{(1)}_{ab} + 8\pi t^{(1)}_{ab} ,$$

where $f_{ab}(g^{(0)}, \mu\gamma^{(L)})$ is linear in $\gamma^{(L)}_{gh}$ and is proportional to μ_{abcdef},

$$T^{(1)}_{ab} \equiv \mathrm{wlim}_{\lambda\to 0}\frac{T_{ab}(\lambda) - T^{(0)}_{ab}}{\lambda} ,$$

and one can write down an explicit formula for $t^{(1)}_{ab}$ in terms of quantities like $\mu^{(1)}_{abcdef}$. It would take several slides to write out the explicit formula for this additional effective source $t^{(1)}_{ab}$ in the general case.

Short Wavelength Deviations

Impose the "wave map gauge" condition

$$\nabla_a H^{ab} = 0$$

where

$$H^{ab} \equiv g^{ab}(0) - \sqrt{\frac{g(\lambda)}{g(0)}} g^{ab}(\lambda) \, .$$

It is convenient to work with $H^{ab}(\lambda)$ rather than $h_{ab}(\lambda)$. The short wavelength part, $H^{ab}_{(S)}(\lambda)$ of $H^{ab}(\lambda)$ does *not* satisfy a linear equation. However, one can write the nonlinear Einstein operator on $H^{ab}_{(S)}(\lambda)$ as the "Lorenz gauge" linearized Einstein operator on $H^{ab}_{(S)}$ plus "pseudotensor" terms, denoted $t'^{ab}(\lambda)$. One can then rewrite this equation as

$$H_S^{\alpha\beta}(x) = 4\int_M G_{\text{ret}}{}^{\alpha\beta}{}_{\mu'\nu'}(x,x')S^{\mu'\nu'}(x') + h_{\text{hom}}^{\alpha\beta}(x)\,,$$

where

$$S^{ab} \equiv T'^{ab}(\lambda) - T^{ab(0)} - \lambda T'^{ab(1)} + t'^{ab}(\lambda) - t^{ab(0)} - \lambda t'^{ab(1)}\,.$$

We argue that under quasi-Newtonian assumptions, $H_{(S)}^{ab}$ is given by a Newtonian approximation, taking only "nearby" matter into account.

Long Wavelength Perturbations in a Locally Newtonian Universe

If we assume that the "short wavelength part" of h_{ab} is described by Newtonian gravity, the formula for $t^{(1)}_{ab}$ simplifies enormously. The combined contribution to the equation for $\gamma^{(L)}_{ab}$ that arises from $t^{(1)}_{ab}$ and "nonlinear corrections" to $T^{(1)}_{ab}$ (i.e., terms that would not arise in ordinary linear perturbation theory) for pressureless matter corresponds exactly to the addition of an effective stress-energy given by the sum of (1) kinetic energy, momentum, and stress terms caused by the motion of the matter and (2) Newtonian gravitational potential energy and stress. As a matter of principle, it is extremely important that these terms are present, but they should be negligible (on all scales at all times) provided only that kinetic energies and binding energies are negligible compared with the perturbed rest mass.

This justifies the use of ordinary linear perturbation theory to describe “long wavelength” phenomena.

Summary

We have developed a new framework/approximation scheme, wherein the "small parameter" is essentially the ratio of the lengthscale of the nonlinearities to the lengthscale of the background curvature. This framework allows $\delta\rho/\rho \gg 1$ and should be applicable to our universe. Our main results for cosmology within this framework are:

- Provided only that the matter always has locally positive energy, the only way small scale inhomogeneities can have a significant effect on large scale FLRW dynamics in our framework is through the presence of gravitational radiation. In particular, small scale matter inhomogeneities can never mimic the effects of dark energy.

- On small scales, the deviation from an FLRW model should be accurately described by local Newtonian gravity.

- Assuming that the gravitational radiation content of our universe is negligible, at long wavelengths, the deviation from an FLRW model should be well described by the quantity $\gamma^{(L)}_{ab}$, which satisfies the ordinary linearized Einstein equation, but has an additional "effective stress-energy source" due to the short wavelength inhomogeneities. It does not appear that these additional terms are of any practical importance.

These results go a long ways toward providing a mathematically consistent justification for the assumptions usually made in cosmology, and provide a framework for improving the approximations.