Consistent cosmological modifications to the Einstein equations

Constantinos Skordis
Perimeter Institute, Waterloo, Ontario N2L 2Y5, Canada
(Dated: June 14, 2008)

General Relativity (GR) is a phenomenologically successful theory that rests on firm foundations, but has not been tested on cosmological scales. The advent of dark energy (and possibly even the requirement of cold dark matter), has increased the need for testing modifications to GR, as the inference of such otherwise undetected fluids, depends crucially on the theory of gravity. In this work I outline a general scheme for constructing consistent and covariant modifications to the Einstein equations. This framework is such that there is a clear connection between the modification and the underlying field content that produces it. I conclude by a simple metric based modification of the fluctuation equations for which the background is exact ΛCDM and present its impact on observations of the cosmic microwave background radiation.

The theory of gravity plays a fundamental role in our modelling and understanding of the universe. If we are to know the matter constituents of the universe, we have to be sure we understand what is the underlying gravitational theory. Einstein’s General Relativity (GR) has played a key role in formulating modern cosmology, first as a smooth Friedmann-Lemaître-Roberson-Walker (FLRW) spacetime, then at the level of linearized fluctuations about this spacetime.

General relativity is a very solid "principle" theory from the theoretical point of view, (and quite understandably the aesthetical point of view). The Lovelock-Grigore theorem [1, 2] asserts that GR with a cosmological constant is unique under the following assumptions: geometry is Riemannian and the gravitational action depends only on the metric, it is local and diffeomorphism invariant and leads to 2nd-order field equations. Relaxing any of these assumptions can lead to more general gravitational theories, e.g. adding extra fields [3, 4, 6], having higher derivatives [7], having a pregeometry [8] or making the theory non-local [9]. This is not an exhaustive list but possible theories fall into one or more categories above.

However as nice as we may think that GR is, the ultimate judge is experiment. Indeed, different aspects of GR have been vigorously tested in the lab, in the solar system and with binary pulsars, all of which lie in the GR have been vigorously tested in the lab, in the solar system and with binary pulsars, all of which lie in the

levels. More recently Hu and Sawicki [13, 14] and Hu [14], have laid down a fully covariant formulation of modifications to gravity under well motivated assumptions.

In this work I outline a general scheme for constructing consistent modifications to the Einstein equations. The scheme is such that, one can clearly classify the modifications according to whether they obey or violate diffeomorphism invariance, need extra fields, or stem from higher derivative theories. Indeed the advantage of this method is the direct connection between the field content and the modifications. Additional assumptions as in [13, 14, 18] can always be used at the very end but we shall not consider this possibility here. I close the paper by constructing the most general modification for which the FLRW background is exact ΛCDM and illustrate the effects on observables in a simple subcase.

Following Hu [14] we start by putting the gravitational field equations in the form

\[ G_{ab} = 8\pi GT_{ab}^{\text{(known)}} + U_{ab} \]  

Here \( G_{ab} \) is the Einstein tensor for the universal matter metric \( g_{ab} \), \( T_{ab}^{\text{(known)}} \) is the stress-energy tensor of all known forms of matter (like baryons, photons and neutrinos), \( U_{ab} \) is a general tensor which encapsulates all the unknown fields/modifications and can depend on \( T_{ab} \) for each field and various combinations of metric functions (such as curvature tensors). The assumption of upto 2nd order field equations, translates to having only upto first derivatives of the extra field in \( U_{ab} \) and upto 2nd derivatives in \( U_{ij} \). Relaxing this assumption is possible and will simply give higher order field equations, but one has to be cautious that quite generally higher derivative theories lead to instabilities.

At this point we have to decide on the field content, i.e. whether \( U_{ab} \) depends on additional fields or the metric alone. In the former case, we must add one field at a time and ensure its energy conservation by applying the Bianchi identity which directly translates to

\[ \nabla_a U^a_{\ b} = 0 \]  

and gives a field equation for the extra field. Violating the Bianchi identity leads to entirely arbitrary parametrizations and will not be considered. The caveat of this approach is that in the case that more than two independent degrees of freedom are present in \( U_{ab} \), one would have to supply extra equations for these, not given by...
the Bianchi identity (see 5, 13, 16 for examples). Finally, there could be interactions between fields in $T_{ab}$ and $U_{ab}$. For simplicity I do not consider this possibility further, but it is straightforward to add.

We now split the dynamics of the problem in the background FLRW dynamics and their fluctuations about that background. The FLRW metric is $ds^2 = a^2(-d\tau^2 + q_{ij}dx^idx^j)$ where $\tau$ is the conformal time, $a$ is the scale factor and $q_{ij}$ is a spacial metric of constant (dimensionless) curvature $K$. The FLRW assumption means that effectively we are considering a collection of scalar fields on the spacial hypersurface of homogeneity and isotropy. This boils down to requiring that the Lie derivative of the extra field vanishes for all six Killing vectors of the FLRW spacetime. Examples are a scalar field $\phi(t)$, a vector field with components $A^\rho = (A(t),0,0,0)$ and a tensor field $X^{\mu\nu}$ whose only non-vanishing components must be $X^{00} = -X(t)$ and $X^{ij} = Y(t)\delta^i_j$. In the special case of a unit-timelike vector field $A(t)$ is pure gauge and the contribution of such a field to the FLRW equations is generally given in terms of functions of $a$, $\dot{a}$ and $\ddot{a}$ [4].

Let's define $E_F = -a^2G^0_0$ and $E_R$ such that $a^2G^0_0 = E_R\delta^i_j$, which are explicitly given by $E_F = 3\frac{\dot{a}}{a} + 3K$ and $E_R = -2\frac{\dot{a}}{a} + a^2 - K$. The FLRW equations corresponding to (1) are then simply written as $E_F = 8\pi Ga^2\sum_i\rho_i + a^2X$ and $E_R = 8\pi Ga^2\sum_i\dot{P}_i + a^2Y$. Applying the Bianchi identity then gives $E_F + \frac{\dot{a}}{a}(E_F + 3E_R) = 0$. and the fluid equation $\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + P) = 0$. Applying the Bianchi identity on $U_{ab}$ imposes

$$\dot{X} + 3\frac{\dot{a}}{a}(X + Y) = 0$$

The above equation will give the background equation for the extra field, or additional constraints on $X$ and $Y$ in the absence thereof. We see that any FLRW background can be modelled via one arbitrary function $Y(t)$ and a second function $X(t)$ found by solving (3).

Adding a scalar field amounts to letting $X = X(\dot{\phi}, \phi, a, \dot{a})$ and $Y = Y(\dot{\phi}, \phi, a, \dot{a}, \ddot{a})$, if we are to expect at most 2nd order field equations for the scalar. For example a canonical scalar field corresponds to $X = \frac{1}{2a^2}\phi^2 + V(\phi)$ and $Y = \frac{1}{2a^2}\phi^2 - V(\phi)$ while the Jordan-Fierz-Brans-Dicke theory [3] to $X = 8\pi Ga^2(e^\phi - 1)\rho + \frac{2}{ae^\phi}\phi^2 + \frac{2\dot{a}}{a}\phi + \frac{3}{2a}\frac{\dot{a}}{a}\phi - \frac{2\dot{a}}{a}\phi + \frac{2\dot{a}}{a}\phi$, and $Y = 8\pi Ga^2(e^\phi - 1)P - \frac{2}{ae^\phi}\phi^2 - \frac{2\dot{a}}{a}\phi + \frac{3}{2a}\frac{\dot{a}}{a}\phi - \frac{2\dot{a}}{a}\phi$.

Consider now the (scalar) fluctuations about the FLRW metric as

$$ds^2 = -a^2(1 - 2\Xi)dt^2 - 2a^2(\nabla_\Xi)dt^2 + a^2[(1 + \frac{1}{3}\chi)q_{ij} + D_{ij}\nu]dx^idx^j$$

where $D_{ij} = \nabla_i\nabla_j - \frac{1}{2}\eta_{ij}\nabla^2$ is a spacial traceless derivative operator. A perfect fluid is described at the fluctuation level by a density contrast $\delta$, momentum $\theta$ such that its total momentum is $u_i = \delta \nabla_i\theta$, dimensionless pressure $\Xi \rightarrow \Xi - \frac{4}{3}$, $\chi \rightarrow \chi + \frac{4}{3}\left[\chi + \frac{2\phi^2}{3}\right]$, $\nu \rightarrow \nu + \frac{2\phi^2}{3}$, $V \rightarrow V + \frac{2\dot{a}}{a}\phi$, $J \rightarrow J + \frac{2\dot{a}}{a}\phi$, $\theta \rightarrow \theta + \frac{1}{a}\xi$, $\theta = \theta + \frac{1}{a}\xi$, $W \rightarrow W + \frac{4}{a}\delta(\frac{a}{3} - \frac{2\dot{a}}{a})\xi$, $\delta \rightarrow \delta + \frac{4}{a}(1 + w)\frac{\dot{a}}{a}\xi + 2\frac{\dot{a}}{a}\phi$, $\delta \rightarrow \delta + \frac{4}{a}(1 + w)\frac{\dot{a}}{a}\xi$, $\delta \rightarrow \delta + \frac{4}{a}(1 + w)\frac{\dot{a}}{a}\xi + 2\frac{\dot{a}}{a}\phi$, $\delta \rightarrow \delta + \frac{4}{a}(1 + w)\frac{\dot{a}}{a}\xi + 2\frac{\dot{a}}{a}\phi$, $\delta \rightarrow \delta + \frac{4}{a}(1 + w)\frac{\dot{a}}{a}\xi + 2\frac{\dot{a}}{a}\phi$, $\delta \rightarrow \delta + \frac{4}{a}(1 + w)\frac{\dot{a}}{a}\xi + 2\frac{\dot{a}}{a}\phi$.

**TABLE I: Gauge transformations for the metric, fluid and Einstein tensor variables.**

If we perform a gauge transformation $X \rightarrow X'$ to all perturbations above, the $\delta G^{0}_0$ Einstein equation becomes

$$-\frac{1}{3}(\nabla^2 + 3K)\left[\chi - \nabla^2\nu'\right] + \frac{\dot{a}}{a}\left[\chi + 2\nabla^2\zeta\right]$$

$$+ \frac{a^2}{2a}\Xi = 8\pi Ga^2\rho \delta$$

i.e. it retains its exact form (it is form-invariant), with the only change being a relabelling of the perturbations with $X \rightarrow X'$ for each variable. Gauge form-invariance always holds for all field equations which stem from a diffeomorphism invariant action, no matter how complicated the theory is.

We can shortcut testing for gauge form-invariance as follows. First define the three gauge non-invariant potentials $V \equiv \nu + 2\zeta$, $J \equiv \chi - \nabla^2\nu$ and $W \equiv \chi + 2\nabla^2\zeta = J + \nabla^2V$, which are the only three combinations of metric variables appearing in the perturbed Einstein tensor. They transform only with the gauge variable $\zeta$. Then define the two gauge-invariant potentials $\Phi \equiv -\frac{1}{6}J + \frac{1}{2}\frac{\dot{a}}{a}V$ and $\Psi \equiv -\Xi - \frac{1}{2}V - \frac{1}{a}\dot{a}V$. We can now split the perturbed Einstein tensor into a gauge invariant and a gauge non-invariant part which involves the variable $V$. For simplicity let us define $E_\Delta = -a^2\delta G^{0}_0$, $E_\theta$ is such that
\[ -a^2 \delta G^0_{i} = \vec{\nabla}^2 E_\theta, \quad E_P = a^2 \delta G^i_{j} \quad \text{and} \quad E_{\Sigma} \quad \text{is such that} \]
\[ a^2 \left[ \delta G^i_{j} - \frac{1}{6} \delta G^k_{k} \delta^i_{j} \right] = D^j E_{\Sigma}. \]
Explicitly we get
\[ E_\Delta = 2(\vec{\nabla}^2 + 3K)\dot{\Phi} - 6\frac{a^2}{a}(\dot{\Phi} + \dot{\tilde{\Psi}}) \]
\[ -3\frac{a}{2} (E_F + E_R)V, \]
\[ E_\Theta = 2(\dot{\Phi} + \dot{\tilde{\Psi}}) + \frac{1}{2}(E_F + E_R)V, \]
\[ E_P = 6 \frac{d}{dt} \left( \frac{\dot{\Phi} + \dot{\tilde{\Psi}}}{a} \right) + 12 \frac{\dot{a}}{a} \frac{\dot{\Phi} + \dot{\tilde{\Psi}}}{a} \]
\[ -2(\vec{\nabla}^2 + 3K)(\dot{\Phi} - \dot{\tilde{\Psi}}) - 3(E_F + E_R)\tilde{\Psi} \]
\[ + \frac{3}{2} \dot{E}_R - 2\frac{\dot{a}}{a} E_R |V, \]
and
\[ E_{\Sigma} = \dot{\Phi} - \dot{\tilde{\Psi}}. \]

The Einstein equations are then given as
\[ E_\Delta = 8\pi Ga^2 \delta \Phi + U_\Delta \]
\[ E_\Theta = 8\pi Ga^2 (\rho + \dot{\rho}) + U_\Theta \]
\[ E_P = 24\pi Ga^2 \rho \Pi + U_P \]
\[ E_{\Sigma} = 8\pi G(\rho + \dot{\rho}) \Sigma + U_{\Sigma} \]
where the \( U_i \) variables are defined in the same way as for the \( E_i \) variables with \( \delta G^0_{i} \) replaced by \( \delta U^0_{i} \).

The importance of gauge form-invariance cannot be overemphasized. In fact gauge form-invariance severely constrains the terms involved. For had we decided to let the RHS of the Einstein equations to be \( 8\pi Ga^2 f_1(\tau) \delta \), \( 8\pi Ga^2 f_2(\tau) \theta \) and \( 24\pi Ga^2 f_3(\tau) \Pi \) we would have found that by virtue of the background equations, the Einstein equations are gauge form-invariant if and only if \( f_1 = f_3 = \rho \), and \( f_2 = \rho + P \). A less trivial example is when \( X = \frac{1}{2} \frac{\dot{a}}{a} \) for some scale \( \dot{r} \). If we decide that this term is due to metric modifications, then \( U_\Delta = \frac{\dot{a}}{2} \left( \frac{\dot{a}}{a} + \frac{\dot{a}}{a} \Sigma \right) \) and \( U_\Theta = -\frac{a^2}{3\dot{a}} \left( \frac{\dot{a}}{a} + \frac{\dot{a}}{a} \Sigma \right) \). If on the other hand we posit that it is due to a unit-timelike vector field perturbed as \( A_i = \frac{\dot{a}}{a} \alpha \) then a possibility is \( U_\Delta = \frac{\dot{a}}{a} (\frac{\dot{a}}{a} - 2\frac{\dot{a}}{a} \alpha) \). In the last case we pick higher derivatives, and a gauge-invariant counter-term must be added to cancel them \[ 17 \].

Once we have added the gauge non-invariant terms and correctly fixed the functions multiplying them by requiring gauge form-invariance to hold, we can proceed to add more gauge-invariant terms involving the extra fields and the gauge invariant potentials \( \Phi \) and \( \tilde{\Psi} \). If we want to consider only parametrizations which lead to 2nd order field equations in all variables, then we have to be careful what terms we add and where. For example, when we add metric terms, we can add up to first derivatives in the two Einstein constraint equations \[ 11 \] and \[ 12 \], and up to second derivatives in the two Einstein propagation equations \[ 13 \] and \[ 14 \]. Since \( \Phi \) is of first order in the metric variables while \( \tilde{\Psi} \) is of second order, we can add \( \Phi \) in all four Einstein equations but \( \tilde{\Psi} \) and \( \Psi \) only in the two propagation equations \[ 13 \] and \[ 14 \]. Note that although \( \Phi + \frac{\dot{a}}{a} \tilde{\Psi} \) is of 1st-order in the metric perturbations, it contains 2nd derivatives of the scale factor, and so it cannot be added to \[ 11 \] and \[ 12 \], unless we relax the 2nd-order field equations constraint. The reason is it is allowed in the definition of \( E_\Delta \) above is because, there are also 2nd derivatives of the scale factor appearing in the gauge non-invariant part of \( E_\Delta \) proportional to \( V \). Thus when \( E_\Delta \) is written in terms of the actual metric potentials, the \( \dot{a} \) terms cancel and the final expression contains only first derivatives in all the variables. On the contrary, when we add solely gauge-invariant terms we no longer have this luxury.

Let us also note that the gauge variable \( \psi \) is not involved in the transformation of the Einstein tensor. Thus if any extra field does transform with \( \psi \) it will always appear in combination with \( \nu, \zeta \) or \( \chi \) in the field equations, in a way that the whole combination does not transform with \( \psi \). An explicit example can be found in \[ 16 \].

We finally utilize the Bianchi identity which at the linearized level simply translates in terms of the added variables \( U_\Delta, U_\Theta, U_P \) and \( U_{\Sigma} \) as
\[ \dot{U}_\Delta + \frac{\dot{a}}{a} U_\Delta - \vec{\nabla}^2 U_\Theta + \frac{1}{2} a^2 (X + Y) W + \frac{\dot{a}}{a} U_P = 0 \] and
\[ \dot{U}_\Theta + 2 \frac{\dot{a}}{a} U_\Theta - \frac{1}{2} U_P - \frac{2}{3} \vec{\nabla}^2 + 3K) U_{\Sigma} + a^2 (X + Y) \Xi = 0 \]
This provides us with the field equations for the extra fields \[ 17 \], or with additional constraints on the added functions in the absence thereof.

I now illustrate the above scheme, by finding the most general diffeomorphism invariant modification to the Einstein equations for which the background cosmology is the plain \( \Lambda \)CDM model, no extra fields are present, and no higher derivative than two is present in the field equations. Since there are no extra fields and the background is unchanged from \( \Lambda \)CDM we can only add gauge invariant terms to Einstein equations by setting \( U_\Delta = A \dot{\Phi}, U_\Theta = B \dot{\Phi}, U_P = C_1 \dot{\Phi} + C_2 \dot{\tilde{\Psi}} + C_3 \dot{\Psi} \) and \( U_{\Sigma} = D_1 \dot{\Phi} + D_2 \dot{\tilde{\Psi}} + D_3 \dot{\Psi}, \) for operators \( A, B, C_i, \) and \( D_i \). Applying the Bianchi identity we get two equations involving \( \dot{\Phi}, \dot{\tilde{\Psi}} \) and \( \dot{\Psi} \) and consistency requires that these equations must be satisfied whatever the values of \( \dot{\Phi}, \dot{\tilde{\Psi}} \) and \( \dot{\Psi} \). A sufficient condition is found by setting the coefficients of these terms to zero which gives \( C_3 = D_3 = 0, \)
\( A = -\frac{2}{3} C_2, B = \frac{1}{3} C_2 + \frac{2}{3} \vec{\nabla}^2 + 3K) D_2 \) and the two differential equations \( A + \frac{2}{3} A - \vec{\nabla}^2 B + \frac{2}{3} C_1 = 0 \) and \( B + 2 \frac{2}{3} B - \frac{2}{3} C_1 - \frac{2}{3} \vec{\nabla}^2 + 3K) D_1 = 0 \). A quick examination reveals that if \( A \) and \( B \) are both zero then we get exact GR. The same holds for \( D_1 \) and \( D_2 \), hence a
The generic prediction of this kind of modification to GR is that $\Phi - \Psi$ should deviate from the GR value.

Let us illustrate the effect on observables in a simple sub-case for which $B = C_1 = D_1 = 0$, $A = \frac{\beta H_0^2}{a}$, $C_2 = -\frac{\beta H_0^2}{a}$, and $D_2 = \frac{3H_0^2}{2a^2}$. Thus, we parametrize deviations from GR with a single dimensionless parameter $\beta$. We let $\gamma = \frac{2\beta H_0^2}{2a^2 + 3H_0^2}$ and find the Einstein equations in the synchronous gauge (defined as $\Xi = \zeta = 0$ and $\chi = h$, $-k^2\nu = h + 6\eta$). The $\dot{h}$ equation is

$$\frac{\dot{a}}{a} = (1 - \gamma)8\pi G a^2 \rho \delta + 2k^2 \eta - 6\gamma \frac{\dot{a}}{a} \dot{\eta} \quad (17)$$

while the $\dot{\eta}$ equation is unchanged.

The perturbation equations were solved numerically in both the synchronous and the conformal Newtonian gauge for numerical consistency. The upper panel of Fig. 1 shows the CMB angular power spectrum $l(l+1)c_l$ for a $\Lambda CDM$ universe ($\beta = 0$) contrasted with non-zero $\beta$. We see that for this particular model, the effect of non-zero $\beta$ is to decrease power on large scales including even the first peak. The lower panel of Fig. 1 shows the time variation of $g = \frac{\dot{h}}{h^2} - \Psi$ for the same set of models at $k = 10^{-3} Mpc^{-1}$. We see that like other modifications to gravity, the effect is to make $g$ grow. In contrast to other parametrizations of modified gravity, however, the difference of $\dot{\Phi} - \dot{\Psi}$ is sourced by $\dot{\Phi}$ rather than $\Phi$.

This scheme can be used to consistently add terms to the Einstein equations which may or may not depend on extra fields, and possible metric functions. The fluctuation equations can then be solved to obtain observable spectra. An important issue is to find predictions concerning these extra parameters for upcoming precision experiments such as the Square Kilometer Array and the EUCLID project.

Acknowledgments: I'm grateful to M. Bañados, R. Durrer, P. Ferreira and W. Hu for important comments. Research at Perimeter Institute for Theoretical Physics is supported in part by the Government of Canada through NSERC and by the Province of Ontario through MRI.