Modifications of gravity

BY CONSTANTINOS SKORDIS*

School of Physics and Astronomy, University of Nottingham, Nottingham NG8 2RD, UK

General relativity (GR) is a phenomenologically successful theory that rests on firm foundations, but has not been tested on cosmological scales. The deep mystery of dark energy (and possibly even the requirement of cold dark matter (CDM)) has increased the need for testing modifications to GR, as the inference of such otherwise undetected fluids depends crucially on the theory of gravity. Here, I discuss 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 argue that this is mandatory for distinguishing modifications of gravity from conventional fluids. I give a non-trivial example, a simple metric-based modification of the fluctuation equations for which the background is exact $\Lambda$CDM, but differs from it in the perturbations. I show how this can be generalized and solved in terms of two arbitrary functions. Finally, I discuss future prospects and directions of research.

Keywords: gravity; cosmology; dark energy

1. Introduction

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 that 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–Lemaitre–Roberson–Walker (FLRW) space–time, then at the level of linearized fluctuations about this space–time.

GR 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 $g_{ab}$, it is local and diffeomorphism invariant and leads to second-order field equations. Relaxing any of these assumptions can lead to more general gravitational theories, e.g. adding extra fields [3–10]; having higher derivatives [11]; having a pregeometry [12–14], additional structure [15,16],

*skordis@nottingham.ac.uk

One contribution of 16 to a Theo Murphy Meeting Issue ‘Testing general relativity with cosmology’.

This journal is © 2011 The Royal Society
a breakdown of geometry [17]; or making the theory non-local [18–20]. This is not an exhaustive list, but possible theories fall into one or more categories mentioned earlier.

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 laboratory, in the Solar System and with binary pulsars, all of which lie in the strong curvature regime (compared with cosmology).

The discovery that the expansion of the Universe is accelerating opens the possibility that GR breaks down on large scales or low curvatures. It may also be that the apparent missing mass in the Universe is not in the form of cold dark matter (CDM), but once again owing to departures from GR. We have the examples of tensor–vector–scalar (T eV eS) [7,21] and Eddington–Born–Infeld [10] theories both of which have been shown to have the potential to do without dark matter and perhaps dark energy (DE) [22–26].

This opens the need for cosmological tests of gravity, and much work has been carried out in this direction [27–44] at various levels.

2. Distinguishing gravity from fluids

We would like to know whether the huge discrepancy between GR on the one hand and the visible (known) matter on the other is due to the breakdown of the former or owing to as yet unknown forms of matter (like dark matter or fluid dark energy).

But at the fundamental level, what is really the distinction? This question has been posed numerous times and the answers vary widely. For example, it has been voiced that \( f(R) \) or scalar tensor theories are not alternative theories of gravity to GR (assertion 1) because their vacuum equations are equivalent to GR coupled to a scalar field, in an appropriate frame (of formulation). According to another point of view, theories with additional fields to the metric, like scalar–tensor or T eV eS theories, are not modifications of gravity (assertion 2). Whether T eV eS is a modification of gravity has also been questioned because it can be written in a single-metric frame [45] where the metric action is Einstein–Hilbert for \( g_{ab} \) while matter follows geodesics of \( g_{ab} \) and does not directly couple to the scalar nor the vector fields (assertion 3). Cosmologically, it has also been argued that if two metric potentials are detected not to be equal, \( \Phi \neq \Psi \), then gravity is modified (assertion 4).

In my opinion, all the earlier-mentioned assertions are incorrect. Assertion 1 does not take into account the way the fields couple to matter. In the Einstein frame (for which the gravitational sector decomposes into the spin-0 and spin-2 parts), the scalar will universally couple to all matter, giving rise to an additional scalar gravitational force. Assertion 2 is clearly misleading: nothing in the notion of what is gravity says anything about the number of mediators or their spins. Assertion 3 is incorrect in a similar way as assertion 1: one has to first write the theory in the diagonal frame [21]; in that case, particles will experience gravitational force owing to all three T eV eS fields. Finally, as to assertion 4, it can be explicitly shown that fluids with shear can easily imitate \( \Phi \neq \Psi \); no-one would argue that photons in the early Universe are a modification of gravity (see [46] for an example regarding the late Universe).
But what is really gravity? It is simply the natural phenomenon of interaction between physical fields with mass (more generally stress–energy). Of course, any field \( U \) has stress–energy and therefore would also contribute to the gravitational field. For example, are the dark fluids (CDM or DE), in this sense, gravitational fields? The key point is that this interpretation must be implemented at the tree level. In that case, we can classify, for our purposes, fields into two classes: gravitational—if they contribute to scattering of point particles in a way that the strength depends on the mass of the particle \( \propto \sqrt{Gm} \), and fluids (e.g. dark matter)—if they do not interact at all at the tree level (but do so at higher loops). Figure 1 schematically depicts this notion.

\[ \text{(a) Parametrizing gravity: a general trick} \]

Even if gravity depends on additional fields or other assumptions, it must always incorporate GR in some limit. Therefore, it must have a metric and this means that it must obey the analogue of the Einstein equations. To deal with a possibly very complicated set of Einstein equations, we perform the following trick [39,42]. We can always rewrite the Einstein equations of any theory such that they look like

\[ G_{ab} = 8\pi G T_{ab} + U_{ab}, \]  

where \( G_{ab} \) is the Einstein tensor of \( g_{ab} \), the metric for which known matter follows geodesics, \( T_{ab} \) is the stress–energy tensor of known forms of matter and \( U_{ab} \) is everything that is left off; it encapsulates the unknown modifications that can be either gravitational or owing to a fluid. Writing the equations this way permits us to use \( \nabla_a G_b^a = 0 \) (Bianchi identity), and \( \nabla_a T_b^a = 0 \) (stress–energy conservation follows from the geodesic assumption). This leads to \( \nabla_a U_b^a = 0 \), which is a set of
field equations for all the modifications (this is a complete set as long as we have only two additional degrees of freedom). Specific examples can be found in the study of Skordis [42].

(b) Impossibility of distinguishing gravity from fluids at the Friedmann–Lemaitre–Roberson–Walker level

Given the earlier-mentioned trick, is it possible to distinguish gravity from fluids at the FLRW level? For FLRW backgrounds, we have that the only degree of freedom in the metric is the scale factor \(a(t)\) such that\( ds^2 = a^2(-dt^2 + dx^2)\), while both \(T_{ab}\) and \(U_{ab}\) can have at most two independent components. In the case of \(T_{ab}\), we have the density \(\rho = -T_{00}\) and pressure \(P = (1/3)T_{ii}\) while we parametrize \(U_{ab}\) as \(U_{00} = -X(t)\) and \(U^i_j = Y(t)\delta^i_j\).

Performing the trick mentioned earlier, we immediately find that

\[
\dot{X} + 3\frac{\dot{a}}{a}(X + Y) = 0. \tag{2.2}
\]

But then \(X\) and \(Y\) cannot be distinguished from \(\rho\) and \(P\) of a conventional fluid. Therefore, we can imitate any expansion history with a fluid, which means that it is impossible to distinguish gravity from fluids at the FLRW level [42].

(c) Parametrizations at the linear level

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

\[
\begin{align*}
    ds^2 &= -a^2(1 - 2\Xi)dt^2 - 2a^2(\nabla_i\xi) dt dx^i + a^2 \left[ \left( 1 + \frac{1}{3}\chi \right) q_{ij} + D_{ij} \right] dx^i dx^j, \tag{2.3}
\end{align*}
\]

where \(D_{ij} \equiv \nabla_i \nabla_j - (1/3)q_{ij}\nabla^2\) is a spatial 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 \(u_i = a\nabla_i\theta\), dimensionless pressure perturbation \(\Pi\) such that \(\delta T_{ij} = \Pi \delta^i_j\) and shear \(\Sigma\) such that the shear tensor is \(\Sigma_{ij} = D_{ij}\).

We then proceed to calculate the perturbed version of equation (2.1), without choosing a gauge yet. The reason is that all sensible theories must lead to perturbed equations that are gauge form-invariant. Gauge transformations are infinitesimal diffeomorphisms generated by a vector field \(\xi^a\) that can be parametrized as \(\xi^a = a(-\xi, \nabla_i\psi)\). Then suppose that a particular perturbed equation can be collectively written as

\[
\sum_A O_A(\tau, x) \Delta_A(\tau, x) = 0, \tag{2.4}
\]

where \(O_A(\tau, x)\) are operators and \(\Delta_A\) are perturbations of various fields. Under gauge transformations, all \(\Delta_A\) would transform to a new \(\Delta'_A\) (although some could be invariant) and the equations would become

\[
\sum_A O_A(\tau, x) \Delta'_A(\tau, x) + [\text{FLRW eq.}]\xi = 0. \tag{2.5}
\]
It is important to realize that the operators $\mathcal{O}_A(\tau, x)$ are unchanged by this transformation. Therefore, the equations retain their form, i.e. they are form-invariant, provided that the background FLRW equations are satisfied. In that case, the only difference between equations (2.4) and (2.5) is a mere relabelling of perturbational variables $\Delta_A \rightarrow \Delta'_A$.

The dependence on the gauge variable $\xi$ can only be eliminated if and only if the background FLRW equations are satisfied. But it turns out that for the whole process to work out, all the gauge-variant terms must be fixed with known functions of the background (see [8] for examples).

(d) Importance of gauge form-invariance and the field content

The majority of the parametrized schemes start by assuming the conformal Newtonian gauge. While these schemes may be consistent, it is far from obvious that they are so. In fact from the way that they are set up, it is impossible to actually test for consistency under gauge form-invariance. Writing the equations in the conformal Newtonian gauge, and then performing a gauge transformation will introduce additional terms that will depend on the gauge variables $\xi$ and $\varphi$. It is not at all clear that the coefficients of the gauge variables will vanish, which is one of the requirements of gauge form-invariance.

The question that arises in the light of the earlier-mentioned statements is whether it might be possible to interpret the potentials appearing in the equations in conformal Newtonian gauge as the gauge-invariant potentials. It might seem that this solves the problem of gauge form-invariance, as all the terms would now be explicitly gauge-invariant. Unfortunately, this interpretation is also incorrect. The reason is that it is impossible, in general, to write down the perturbed field equations (whether Einstein equations or any other set of field equations), such that all terms that appear are by construction gauge-invariant. The only case that this is actually possible is when the background tensors are constant, which is forbidden in the case of an FLRW Universe. This is a consequence of the well-known Stewart–Walker lemma [47]. In other words, although it is possible to write any perturbed field equation as a sum of gauge-invariant terms, each term cannot, in general, arise as a perturbation of a tensor constructed out of the fields of the theory. Such a construction is nothing more than a convenient mathematical construct, but otherwise physically empty.

The true power of gauge form-invariance manifests in conjunction with the specification of the field content of the parametrization. Different fields transform differently under gauge transformations, and this is then directly linked to the individual gauge-variant terms that can be a part of $U_{b}^{a}$. However, it might not be directly obvious why specifying the field content of $U_{b}^{a}$ is by itself important. After all, why not simply consider the gauge transformation of the whole of $U_{b}^{a}$ and ignore its composition?

Let us exemplify. The tensor $U_{b}^{a}$ transforms like a stress–energy tensor, and this is enough for constructing consistent parametrizations of $U_{b}^{a}$. Such is the approach followed in Hu [39]. However, this prohibits a direct physical interpretation of our findings in the case that a non-zero contribution from $U_{b}^{a}$ is detected. More specifically, it is impossible to attribute this contribution to a modification of gravity as opposed to the presence of some ordinary unknown fluid without further assumptions. For example, the difference $\Phi - \Psi$, commonly
(and incorrectly) thought of as indicating a modification of gravity, could also be sourced by a standard fluid with shear as has been mooted by Kunz & Sapone [46] and Bertschinger & Zukin [41]. Specifying the field content of $U^a_b$ is the extra assumption that we need to distinguish between a modification of gravity and gravitational effect of standard fluids. Once the fields comprising $U^a_b$ are specified, we can proceed to answer the question: ‘What is the force between two well-separated masses in vacuum?’ We can then distinguish gravity from fluids depending on whether the field equations lead to a modification of the standard gravitational law or not. I shall not consider the details of how this last step is performed (the reader is referred to Iliopoulos et al. [48] for the case of GR in an expanding background) in this work, but only consider the way of how such field equations can be consistently written.

3. Simple examples with up to two time derivatives

(a) The extended $\Lambda$ cold dark matter model

I further illustrate the earlier-mentioned scheme with a less trivial example than the conventional fluid. In what follows I find 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 add only gauge-invariant terms to the Einstein equations. We have at our disposal two gauge-invariant combinations of the metric perturbations, namely

$$\hat{\Phi} = \frac{1}{6}(-\chi + \nabla^2 \nu) + \mathcal{H} \left( \frac{1}{2} \dot{\nu} + \zeta \right)$$  \hspace{1cm} (3.1)

and

$$\hat{\Psi} = -\Xi - \frac{1}{2} \ddot{\nu} - \dot{\zeta} - \mathcal{H} \left( \frac{1}{2} \dot{\nu} + \zeta \right).$$  \hspace{1cm} (3.2)

We first let $U_\Delta = -a^2 \delta U^0_0$, $U_\Theta$ such that $-a^2 \delta U^0_i = \nabla_i U_\Theta$, $U_P$ by $U_P = a^2 \delta U^i_i$ and $U_\Sigma$ as $a^2 [\delta U^i_j - (1/3) \delta U^k_k \delta^i_j] = D^i_j U_\Sigma$. The only allowed gauge-invariant modifications are $U_\Delta = A \hat{\Phi}$, $U_\Theta = B \hat{\Phi}$, $U_P = C_1 \hat{\Phi} + C_2 \hat{\Phi} + C_3 \hat{\Psi}$ and $U_\Sigma = D_1 \hat{\Phi} + D_2 \hat{\Phi} + D_3 \hat{\Psi}$, for spatial pseudo-differential operators $A, B, C, D$ and $D$. Applying the Bianchi identity, we get two equations involving $\hat{\Phi}$, $\hat{\Psi}$ and consistency requires that these equations must be satisfied whatever the values of $\hat{\Phi}$, $\hat{\Phi}$ and $\hat{\Psi}$. A sufficient condition is found by setting the coefficients of these terms to zero. This gives $C_3 = D_3 = 0$, the two constraints

$$A = -\mathcal{H} C_2,$$  \hspace{1cm} (3.3)

$$B - \frac{1}{3} C_2 = \frac{2}{3} (\nabla^2 + 3K) D_2$$  \hspace{1cm} (3.4)
and the two differential equations
\[ \dot{A} + \mathcal{H}(A + C_1) - \nabla^2 B = 0, \]
\[ \dot{B} + 2\mathcal{H}B - \frac{1}{3}C_1 - \frac{2}{3}(\nabla^2 + 3K)D_1 = 0. \]  

A quick examination reveals that if \( A \) and \( B \) are both zero, then we get exact GR. The same holds if \( D_1 \) and \( D_2 \) are also both zero; hence a generic prediction of this kind of modification to GR is that \( \dot{\Phi} - \dot{\Psi} \) should deviate from the GR value. Yet another special case is when \( D_2 = B = 0 \), but \( D_1 \) is not assumed at first to vanish. Using the earlier-mentioned conditions, however, we find that all operators must vanish and once again we recover GR.

To illustrate the effect on observables, let us make further assumptions and consider a simple subcase for which the space–time is spatially flat and for which \( B = C_1 = D_1 = 0 \). The only non-zero operators are \( A = \beta H_0^2/a \), \( C_2 = -\beta H_0^2/\dot{a} \) and \( D_2 = (\beta H_0^2/2\dot{a})(1/\nabla^2) \). Thus, we parametrize deviations from GR with a single dimensionless parameter \( \beta \) that appears only in the perturbed equations and not in the background. The action of \( 1/\nabla^2 \) is defined by its spectral representation. Given a pseudo-differential operator \( D(t) \) acting on \( F(t, x) \), we have
\[ D(t)F(t, x) = \int \frac{d^3k}{(2\pi)^3} e^{ikx} d(t, k)f(t, k). \]

The function \( d(t, x) \) is called the symbol. The symbol for \( \nabla^{-2} \) is \( -k^{-2} \). More generally, we will use symbols that are dimensionless, by an appropriate rescaling with factors of \( k \), i.e. \( A \leftrightarrow k^2 A \), \( B \leftrightarrow kB \), \( C_1 \leftrightarrow k^2 C_1 \), \( C_2 \leftrightarrow kC_2 \), \( D_1 \leftrightarrow D_1 \) and \( D_2 \leftrightarrow D_2/k \).

How does this modification fit in accordance with the Lovelock–Grigore theorem? It is clear that the operators above contain inverse powers of \( \dot{a} \) as well as the pseudo-differential operator \( 1/\nabla^2 \). This means that if a full nonlinear theory exists that leads to an exact \( \Lambda \)CDM background and that deviates at the perturbative level as earlier mentioned, then such a theory must be non-local. This is in full accordance with the Lovelock–Grigore theorem.

At this point, we can finally pick a gauge of choice and solve the field equations. I have solved the equations for the above model numerically both in the synchronous and in the conformal Newtonian gauge for numerical consistency (see [42]). To solve the perturbed equations, we must also specify the initial conditions and I further assume that the initial conditions are adiabatic. This introduces a \( \beta \) dependence in the adiabatic-growing mode to order \( k\tau \) in the synchronous gauge (which vanishes for \( \beta = 0 \)), while in the conformal Newtonian gauge there is no such dependence to leading order in \( k\tau \), but it arises at higher orders. Figure 2a shows the cosmic microwave background (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. Figure 2b shows the time variation of \( \dot{\Phi} - \dot{\Psi} \) for the same set of models at \( k = 10^{-3}\text{Mpc}^{-1} \). We see that like other modifications to gravity, the effect is to make \( \dot{\Phi} - \dot{\Psi} \) grow. In contrast to conventional parametrizations of modified gravity [29,49] however, the difference of \( \dot{\Phi} - \dot{\Psi} \) is sourced by \( \dot{\Phi} \) rather than \( \Phi \).
The simple example provided earlier, where $D_2 = (\beta H_0^2/2\dot{a})(1/\nabla^2)$, can easily be generalized. We may solve the operator equations (3.3)–(3.6) analytically [44]. The solution is in terms of two free functions: $D_1$ and $D_2$. We get

\begin{align}
A &= 2\mathcal{H}\frac{\mathcal{H}_k\dot{D}_2 + k(2\mathcal{H}_k^2 + 1/3)D_2 - \mathcal{H}D_1}{\mathcal{H} - \mathcal{H}^2 - k^2/3}, \\
B &= -\frac{k}{3\mathcal{H}}A - \frac{2}{3}D_2, \\
C_1 &= \frac{3}{k}(\dot{B} + 2\mathcal{H}B) + 2D_1 = -A - \frac{1}{\mathcal{H}}[\dot{A} + kB] \\
C_2 &= -\frac{k}{\mathcal{H}}A,
\end{align}

(3.8)

where I have defined $\mathcal{H}_k = \mathcal{H}/k$. From now on, I choose the conformal Newtonian gauge defined by $\Psi = -\tilde{\zeta}$, $\zeta = v = 0$ and $\Phi = -(1/6)\chi$. With this in hand, we find that the potentials $\Phi$ and $\Psi$ (not to be confused with the gauge-invariant potentials $\hat{\Phi}$ and $\hat{\Psi}$) evolve as

\begin{equation}
\Phi = -\frac{3\mathcal{H}_k^2 \sum X \Omega_X [\Delta_X + 3(1 + w_X)\mathcal{H}\theta_X]}{2(1 - \mathcal{H}_kD_2) + 9\mathcal{H}_k^2 \sum X (1 + w_X)\Omega_X}
\end{equation}

(3.9)
and

$$\psi = \frac{[1 - D_1 - (B/2)D_2] \Phi - (3/2)\mathcal{H}_k^2 \sum_X \Omega_X (1 + w_X) (kD_2 \theta_X + 3\sigma_X)}{1 - \mathcal{H}_k D_2}, \quad (3.10)$$

where $\Delta_X \equiv \delta_X - 3\mathcal{H}(1 + w_X)\Phi$, and $\sigma_X$ is the fluid shear.

(b) Generalizing the model

It turns out that we can relax the condition that the background is the $\Lambda$CDM model. In particular, we assume that the model contains a DE fluid with a general time-dependent equation of state $w = \bar{P}/\bar{\rho}$ (where $\bar{\rho}$ and $\bar{P}$ are the background DE density and pressure, respectively). The perturbation to the DE stress–energy tensor is parametrized as $\delta T^0_0 = -\bar{\rho} \delta E, \delta T^i_i = -(\bar{\rho} + \bar{P}) \nabla \theta_E, \delta T^i_j = \bar{\rho} \Pi_E \delta^i_j$, where $\delta E, \theta_E$ and $\Pi_E$ are the DE density contrast, momentum divergence and pressure contrast, respectively. We assume that DE has an arbitrary speed of sound $c^2_s$ but no shear. The speed of sound $c^2_s$ is then defined via the equation

$$\Pi_E = c^2_s \delta E + 3(c^2_s - c^2_a)(1 + w_E)\mathcal{H} \theta_E, \quad (3.11)$$

where $c^2_a \equiv w - \dot{w}/3\mathcal{H}(1 + w)$ is the adiabatic speed of sound.

We further assume that $\nabla^a T^\text{CDM}_{ab} = \nabla^a T^\text{DE}_{ab} = 0$, which imposes $\nabla^a U_{ab} = 0$ and that, as before, $U_{ab}$ contains only metric modifications. Clearly, then, $\delta U_{ab}$ is gauge-invariant and in fact one gets the same set of conditions (3.3), (3.4), (3.5) and (3.6) on the operators as in the $\Lambda$CDM case. The solution is once again (3.8) and the Newtonian gauge potentials evolve as equations (3.9) and (3.10); what is new here is that the $\sum_X$ in equations (3.9) and (3.10) runs over the DE as well.

4. Derived parametrizations

The equations (3.9) and (3.10) can be solved along with the evolution equations for the fluid perturbations, given a general set of initial conditions and two general functions $D_1(t, k)$ and $D_2(t, x)$. However, it is possible under certain assumptions to reduce the functional form of $D_1(t, k)$ and $D_2(t, x)$ into a set of constants and determine the effect on observables when those constants are varied.

(a) Parametrizing the modified gravity operators

In the parametrized post-Newtonian (PPN) approach to solar system tests, one isolates the solutions of the potentials from the parameters. We would like to invoke a similar approach here. In what follows, I assume that the modification of gravity is due to the presence of DE. This means that $D_1$ and $D_2$ are expected to vanish in the limit $\Omega_E \to 0$. I will further assume that on small scales, i.e. in the limit $\mathcal{H}_k \to 0$, the perturbed Einstein equations are well behaved and, moreover, they reduce to the PPN formalism to linear order. Thus, we expand $D_1$ and $D_2$ in a Taylor series in $\mathcal{H}_k$ and $\Omega_E$:

$$D_1 = \zeta_1 \Omega_E + \zeta_2 \Omega_E^2 + \cdots + O(\mathcal{H}_k) \quad (4.1)$$
and

\[ \mathcal{H}_k D_2 = g_1 \Omega_E + g_2 \Omega_E^2 + \cdots + O(\mathcal{H}_k). \]  

(4.2)

Note how the expansion concerns \( \mathcal{H}_k D_2 \) rather than \( D_2 \). This is imposed by the requirement that the PPN limit on small scales is reachable (see [44]). This way of parametrizing \( D_1 \) and \( D_2 \) has three major advantages.

— It is in the spirit of the PPN formalism, where the PPN parameters are isolated from the potentials that are dependent on the density profiles and thus the solutions; the role of the ‘potential’ in this case is taken by \( \Omega_E(\tau) \), which depends on the background cosmology.

— Expanding in powers of \( \Omega_E \) isolates the background effects of the DE from the genuine effects of the perturbations. In particular, the DE relative density \( \Omega_{0E} \) or the dark matter relative density \( \Omega_{0m} \) would have no effect on the parameters \( \xi_i \) and \( g_i \).

— This expansion makes mathematical sense for any analytical function, as the function \( \Omega_E(\tau) \) is always bounded to be \( 0 \leq \Omega_E \leq 1 \), i.e. it is a naturally small parameter.

The constants \( \xi_i \) and \( g_i \) are to be determined from observations. The effect of varying \( \xi_1 \) and \( g_1 \) on the CMB angular power spectrum and the matter power spectrum is shown in Figure 3. Investigations are currently underway to impose constraints on these parameters from observations of the CMB and large-scale structure.

(b) Parametrizing the growth rate

Peebles [50] proposed that the logarithmic derivative of the growth rate may be given by

\[ f = \frac{d \ln \delta_M}{d \ln a} = \Omega_M^\gamma \]  

(4.3)

(subsequently rederived and advocated in [51] and [52]), where \( \Omega_M \) is the matter density and \( \gamma \approx 0.6 \) for \( \Lambda \text{CDM} \) is a parameter. It has been shown that models different from \( \Lambda \text{CDM} \) lead to a different value for \( \gamma \) and if this could be detected, then it would be a nice model discriminator. I will discuss shortly how we can determine \( \gamma \) in the model discussed earlier.

Given our parametrization of \( D_1 \) and \( D_2 \), we can impose a similar parametrization for \( w \). We let

\[ w_E = w_0 + w_1 \Omega_E + w_2 \Omega_E^2 + \cdots \]  

(4.3)

We will, in addition, be looking at late times where the only fluids in the Universe are effectively CDM and DE. In that case, \( \Omega_E \) evolves as \( \dot{\Omega}_E = -3 \mathcal{H} w \Omega_M \Omega_E \) and in addition we have \( \dot{\mathcal{H}} = -(1/2) \mathcal{H}^2 (1 + 3 w \Omega_E) \). Assuming once again small scales, we find [44] that

\[ \gamma = \gamma_0 + \gamma_1 \Omega_E + \gamma_2 \Omega_E^2 + \cdots, \]  

(4.4)

where \( \gamma_i \) are constants that are model-dependent. For instance, \( \Lambda \text{CDM} \) gives \( \gamma_i = \{ 6/11, 15/2057, 4205/1040842 \ldots \} \) and Dvali–Gabadadze–Porrati (DGP) gives \( \gamma_i = \{ 11/16, 7/5632, -93/4096 \} \). In general, \( \gamma_i \) is a function of \( w_i \), \( g_i \) and \( \xi_i \). The expressions are lengthy and can be found in Ferreira & Skordis [44]. This approximation can be shown to be better and a few per cent even for values as high as \( \Omega_E \sim 0.9 \) [44].
5. Further developments

(a) Parametrizing theories with arbitrary dynamical fields

The condition $\nabla^a T_{ab}^{\text{DE}} = 0$ need not be valid in general, but rather $\nabla^a (T_{ab}^{\text{DE}} + U_{ab}) = 0$. Such is the case of DGP or generalized Einstein aether. In such a case, it is difficult to decide what we call $\delta T_{ab}$ and what we call $U_{ab}$, because in general such a $\delta T_{ab}$ cannot be put in the generalized fluid form. We can make further progress, however, by specifying the field content and construct the general theory, given that field content. This brings us back to what I discussed in the beginning: that distinguishing the effects of gravity from the effects of fluids requires us to specify the field content.

As an example, consider the case where our field content, in addition to the known matter fields, is a unit-time-like vector field $A_a$. Such a vector field would have a scalar mode $\alpha$ given by $A_i = a \nabla_i \alpha$. If we require that we have up to two time derivatives in the field equations, we can write

$$U_\Delta = -\frac{1}{2} a^2 (X + Y)(\chi - k^2 \nu) + A_1 \dot{\phi} + A_2 \dot{\alpha} + A_3 \dot{\beta},$$

where $\dot{\alpha} = \alpha - (1/6 \mathcal{H})(\chi - k^2 \nu)$ and $\dot{\beta} = \dot{\alpha} + \dot{\Xi} + (1/6)(\chi - k^2 \nu)$ are two gauge-invariant variables which involve the vector field. Similar expressions are found for the other $U$’s. Gauge transforming and requiring that the field equations are
gauge form-invariant, we find a set of conditions. When re-inserted back into
the expressions for $U_{ab}$ and after the Bianchi identity is used we get two field
equations for $\alpha$ of the form

$$\mathcal{O}^{(i)}_{1} \ddot{\alpha} + \mathcal{O}^{(i)}_{2} \dot{\alpha} + \mathcal{O}^{(i)}_{3} \dot{\Phi} + \mathcal{O}^{(i)}_{4} \dddot{\psi} + \mathcal{O}^{(i)}_{5} \dot{\Phi} + \mathcal{O}^{(i)}_{6} \dddot{\psi} = 0, \quad i = 1 \ldots 2, \quad (5.2)$$

where $\mathcal{O}^{(i)}_{i}$ are a set of operators of background functions. The solution has to be
consistent with both equations; hence, there exists a set of conditions that $\mathcal{O}^{(i)}_{i}$
have to satisfy.

(b) Parametrizing higher derivative theories

We may relax the requirement that the tested theory has up to two time
derivatives. Doing so requires us to use more operators. For example, if we
consider theories with at most $N$ time derivatives, we let

$$U_{\Delta} = \sum_{i=0}^{N-2} \mathcal{A}^{(\Phi)}_{i}(t) \frac{d^{i}}{dt^{i}} \Phi + \sum_{i=0}^{N-3} \mathcal{A}^{(\psi)}_{i}(t) \frac{d^{i}}{dt^{i}} \dddot{\psi}, \quad (5.3)$$

and similarly for the other $U$’s. If in addition, $\nabla_{a} T^{\text{DE}}_{ab} = 0$, the Bianchi identity
imposes a series of conditions on the operators. In fact $f(R)$ theories fall under
this category with $N = 4$. I do not discuss these possibilities further and refer to
work in progress by Baker et al. [53].

6. Conclusions

I have presented a scheme that prescribes how consistent modifications of
the Einstein equations can be constructed. At the heart of the scheme lies
the physical requirement that the linearized field equations of any theory should
be gauge form-invariant. This requires the specification of the field content of the
theory and thus provides the means for distinguishing modifications of gravity
from effects coming from conventional matter fluids. The resulting fluctuation
equations can then be solved to obtain observable spectra on the scales of interest
for any set of initial conditions. Future work would include more refinements with
focus on current and future observational constraints.

Part of this work was done in collaboration with T. Baker, P. Ferreira and J. Zuntz.

References

1 Lovelock, D. 1969 The uniqueness of the Einstein field equations in a four dimensional space.
Quan. Grav. 9, 1555. (doi:10.1088/0264-9381/9/6/012)
(doi:10.1038/164637a0)
4 Fierz, M. 1956 On the physical interpretation of P. Jordan’s extended theory of gravitation.
Rev. 124, 925–935. (doi:10.1103/PhysRev.124.925)
Modifications of gravity


