Cosmological probes of modified gravity: the nonlinear regime

BY FABIAN SCHMIDT*

Theoretical Astrophysics, California Institute of Technology, Mail Code 350-17, Pasadena, CA 91125, USA

We review the effects of modified gravity on large-scale structure in the nonlinear regime, focusing on $f(R)$ gravity and the Dvali–Gabadadze–Porrati model, for which full $N$-body simulations have been performed. In particular, we discuss the abundance of massive halos, the nonlinear matter power spectrum and the dynamics within clusters and galaxies, with particular emphasis on the screening mechanisms present in these models.

Keywords: cosmology; theory; modified gravity; large-scale structure; gravitational lensing

1. Introduction

The regime of nonlinear structure formation, i.e. on scales of tens of Mpc or smaller, offers very rich possibilities to search for and constrain modifications to gravity. Any viable gravity model that deviates significantly from general relativity (GR) on large scales needs to invoke a nonlinear mechanism to restore GR on small scales. This is necessary as solar system tests put tight constraints on deviations from GR; in particular, light universally coupled scalar fields. In this review, we focus on models that, at least on the scales of interest, can be described as scalar–tensor theories. The majority of modified gravity scenarios discussed in the cosmological context fall into this class, with the notable exception of tensor–vector–scalar theory [1]. Moreover, the main motivation for the models we will consider is to explain the late-time acceleration of the universe, and they employ cold dark matter in the same way as the standard Lambda cold dark matter (LCDM) scenario. Once the matter density field becomes nonlinear, these ‘screening mechanisms’ can—in general—no longer be neglected, and need to be included either via full simulations or semi-analytical modelling. Thus, strictly speaking, a rigorous model-independent parametrization of modified gravity such as what has been performed in linear cosmological perturbation theory [2–4] or in the solar system (parameterized post-Newtonian formalism) is not possible in the nonlinear regime.

*fabians@caltech.edu

One contribution of 16 to a Theo Murphy Meeting Issue ‘Testing general relativity with cosmology’.
Several nonlinear mechanisms that hide scalar fields in the solar system have been proposed in the literature: the chameleon [5,6] and symmetron [7] mechanisms invoke suitable scalar-field potentials that allow the scalar degree of freedom to become very massive in high-density environments. Galileon fields [8] inspired by braneworld scenarios [9] have strong derivative self-interactions that lead to a decoupling in dense environments (Vainshtein mechanism [10,11]).

In the following, we will concentrate on two specific scenarios, which will be introduced in §2: $f(R)$ gravity and the Dvali–Gabadadze–Porrati (DGP) braneworld model. The main reason for this choice is that self-consistent $N$-body simulations of structure formation have been performed in these models, so that the effect on nonlinear structure can be studied quantitatively. Furthermore, each model employs a different type of nonlinear screening mechanism, and the results qualitatively represent the ‘chameleon’ and ‘galileon’ classes of models.

After introducing the modified gravity models in §2, we consider the mass function of dark matter halos and their clustering in §3. We then move on to the matter power spectrum and weak lensing in §4. Finally, we discuss the dynamics of matter within virialized structures in §5. We conclude in §6.

2. Models

(a) $f(R)$ gravity

In the $f(R)$ model [12–14], the Einstein–Hilbert action is augmented with a general function of the scalar curvature $R$,

$$S_G = \int d^4x \sqrt{-g} \left[ \frac{R + f(R)}{16\pi G} \right], \quad (2.1)$$

where $f(R)$ gravity is equivalent to a scalar–tensor theory [15], and $f_R = df/dR$ is the additional dimensionless scalar degree of freedom. This field has a mass and propagates on scales smaller than the associated Compton wavelength $\lambda_C = \sqrt{3df_R/dR}$. Well within the Compton wavelength, the scalar mediates a $4/3$ enhancement of gravitational forces, with corresponding strong effects on the growth of structure in the universe.

While these general features are common to all viable functional forms of $f(R)$, to make progress, we will consider the model of Hu & Sawicki [16],

$$f(R) = -2\Lambda \frac{R^n}{R^n + \mu^2}, \quad (2.2)$$

with three free parameters, $n$, $\Lambda$ and $\mu^2$. Note that as $R \to 0$, $f(R) \to 0$, and hence the model does not contain a cosmological constant. Nevertheless, for $R \gg \mu^2$, the function $f(R)$ can be approximated as

$$f(R) = -2\Lambda - \frac{f_{R0}}{n} \frac{\bar{R}_0^{n+1}}{R^n}, \quad (2.3)$$

with $f_{R0} = -2\Lambda \mu^2 / \bar{R}_0^2$ replacing $\mu$ as the second parameter of the model. Here, we define $\bar{R}_0 = \bar{R}(z = 0)$, so that $f_{R0} = f_R(\bar{R}_0)$, where overbars denote the quantities.
of the background spacetime. Note that if $|f_{R0}| \ll 1$, then the curvature scales set by $A = O(R_0)$ and $\mu^2$ differ widely and hence the $R \gg \mu^2$ approximation is valid today and for all times in the past.

The background expansion history mimics ΛCDM with $A$ as a true cosmological constant to order $f_{R0}$. Therefore, in the limit $|f_{R0}| \ll 10^{-2}$, the $f(R)$ model and ΛCDM are essentially indistinguishable with geometrical tests. Nonetheless, geometrical tests do constrain one of the two parameters ($A$), leaving large-scale structure to constrain the other parameter ($f_{R0}$) that controls the strength and range of the force modification. The parameter $n$ controls how rapidly the field amplitude evolves over cosmic time. Later, we will consider only the case $n = 1$.

On subhorizon scales, where time derivatives may be neglected with respect to spatial derivatives, the equation for $f_R$ field perturbations $\delta f_R = f_R - f(\bar{R})$ is given by

$$\nabla^2 \delta f_R = \frac{a^2}{3} [\delta R(f_R) - 8\pi G\delta \rho_m], \quad (2.4)$$

where $\delta R(f_R)$ encodes the nonlinearity. The $\delta f_R$ field fluctuation then acts as an additional source to the gravitational potential (i.e. the perturbation to the time–time component of the metric in longitudinal gauge),

$$\nabla^2 \Psi = 4\pi G a^2 \delta \rho_m - \frac{1}{2} \nabla^2 \delta f_R. \quad (2.5)$$

In the regime, where $f(\bar{R})$ is larger than typical gravitational potential wells, perturbations in the $f_R$ field are small, so that $\delta R \approx 0$ in equation (2.4). Through equation (2.5), gravitational forces are enhanced by 4/3. If $f(\bar{R})$ is of the order of the gravitational potential or smaller, then the r.h.s. of equation (2.4) must adjust to be close to 0, and hence GR is restored.

In the study of Oyaizu and co-workers [17,18], N-body simulations were performed that solve the system equations (2.4) and (2.5). High-resolution simulations of the same model have recently been carried out [19]; their results are generally in agreement with the findings reported here. In addition to the full $f(R)$ simulations and ΛCDM simulations using the same initial conditions, simulations were performed using the linearized version of equation (2.4). In these ‘no-chameleon’ simulations, gravity is enhanced within $\lambda_C$, regardless of the local potential well. In order to gain qualitative insight into the effects of $f(R)$ on structure formation, we will also discuss the predictions of two limiting cases of the collapse of a spherical tophat perturbation. The first case corresponds to ordinary gravity, which applies if either the perturbation is larger than the Compton wavelength at all times, or is chameleon screened. The second case assumes forces enhanced by 4/3 throughout.

(b) The Dvali–Gabadadze–Porrati model

The DGP model [9] consists of a spatially three-dimensional brane in a 4 + 1 dimensional Minkowski bulk. Matter and all interactions except gravity are confined to the brane. The gravitational action consists of the five-dimensional Einstein–Hilbert action plus a four-dimensional Einstein–Hilbert term that leads to the recovery of four-dimensional gravity on small scales. The two gravitational constants, or Planck masses $M_{Pl(5)}$, $M_{Pl(4)}$ appearing in the action can be related
via the crossover scale \( r_c \equiv M_{\text{Pl}(4)}^2 / 2M_{\text{Pl}(5)}^3 \). On scales above \( r_c \), gravity becomes five dimensional, with forces falling off as \( 1/r^3 \). The Friedmann equation for the scale factor on the brane is modified in the DGP model as follows [20]:

\[
H^2 \pm H \frac{r_c}{r} = \frac{8\pi G}{3} \bar{\rho},
\]

(2.6)

where \( \bar{\rho} \) is the total average energy density on the brane including any dark energy or brane tension. The sign choice on the l.h.s. corresponds to the choice of branch of the model, ‘−’ leads to the self-accelerating branch, whereas ‘+’ corresponds to the normal branch.

On scales much smaller than both the horizon \( H^{-1} \) and the cross-over scale \( r_c \), the DGP model reduces to an effective scalar–tensor theory with the dimensionless brane-bending mode \( \varphi \) representing the scalar. Well within the horizon, time derivatives can be neglected with respect to spatial derivatives and the equation for the brane-bending mode can be written as [21,22]

\[
\nabla^2 \varphi + \frac{r_c^2}{3\beta} [(\nabla^2 \varphi)^2 - (\nabla_i \nabla_j \varphi)(\nabla^i \nabla^j \varphi)] = \frac{8\pi G}{3\beta} \delta \rho,
\]

(2.7)

where the function \( \beta(a) \) is given by

\[
\beta(a) = 1 \pm 2H(a) r_c \left( 1 + \frac{\dot{H}(a)}{3H^2(a)} \right).
\]

(2.8)

Here, again, the positive sign holds for the normal branch of the DGP model, whereas the negative sign holds for the self-accelerating branch. The brane-bending mode then couples to the potential through equation (2.5) just like in the \( f(R) \) case, with \( \delta f_R \rightarrow \varphi \).

In the study of Schmidt [23,24], N-body simulations were performed that solve equation (2.7) for the self-accelerating and normal branches, respectively. In addition to the full simulations and those of a GR model with the same expansion history, simulations were performed that used a linearized version of equation (2.7) (‘linearized DGP’). In the following, we will show results from the self-accelerating branch (sDGP), though the normal branch behaves qualitatively similar (with opposite sign).

The collapse of a tophat perturbation in the DGP model has been studied earlier [25,26]; interestingly, the tophat remains a tophat during collapse, and the collapse dynamics depend only upon the collapse redshift and \( r_c \), not on the scale of the tophat. We will compare the results based on these collapse calculations with the full simulation results in the following sections.

3. Halo abundance

The abundance of galaxy clusters, which are associated with rare, massive halos, is a sensitive probe of the history of structure formation, and has already successfully been used as a probe of gravity [27–32]. The enhancement of the halo abundance measured in \( f(R) \) (figure 1) and DGP (figure 2) simulations is qualitatively what one would expect from an enhanced (or suppressed) linear

\[
\text{Phil. Trans. R. Soc. A} (2011)
\]
growth. The shaded band in each figure shows the results of the spherical collapse prediction combined with the Sheth–Tormen mass function prescription [33]. The effect of modified gravity is generally most evident at the high-mass end. However, the most massive objects are also generally most susceptible to the nonlinear screening mechanisms. This is particularly evident in the $f(R)$ case, where the abundance enhancement for the most massive clusters is shut off at the high mass end for $|f_{R0}| \lesssim 10^{-5}$. For small-field values ($|f_{R0}| \sim 10^{-6}$), the abundance of massive clusters is not sensitive to the $f_R$ field anymore, and galaxy groups rather than clusters become a more effective probe. Note that in this regime, the simple spherical collapse model is not a good description anymore, while it provides a good description of the strong field case ($|f_{R0}| > 10^{-5}$).

As the Vainshtein mechanism in the DGP model operates above a certain density threshold rather than a certain depth of the potential well, its impact on the mass function is almost mass independent, as can be seen by comparing the full and linearized DGP results in figure 2. This feature is reproduced in the spherical collapse model, while the model does not quite reproduce the amplitude of the suppression.

The observed abundance of X-ray and optically selected clusters has been used to obtain the tightest constraints on the $f(R)$ model outside the solar system [30,32,34]. Figure 3 shows the constraints on the field amplitude and Compton wavelength (reach) of the $f_R$ field as a function of the model parameter $n$. Clearly, a significant portion of parameter space can already be ruled out, with

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1}
\caption{Relative deviations of the $f(R)$ halo mass functions from $\Lambda$CDM, with (a) $|f_{R0}| = 10^{-4}$, (b) $10^{-5}$ and (c) $10^{-6}$. In each case, squares denote the full simulations, whereas triangles (displaced horizontally for visibility) denote the no-chameleon simulations. The shaded band shows the range of enhancement expected from spherical collapse. Reproduced with permission from Schmidt et al. [33]. (Online version in colour.)}
\end{figure}
Figure 2. Deviation in the halo mass function at $z = 0$ of the self-accelerating DGP model from a dark energy model with the same expansion history (here referred to as QCDM). The points show measurements in the full (squares) and linearized DGP (triangles) simulations, the shaded band shows the Sheth–Tormen + spherical collapse prediction range between full DGP collapse (dashed lines) and linearized DGP collapse. Reproduced with permission from Schmidt et al. [26]. (Online version in colour.)

Figure 3. (a) Constraints on $|f_{R0}|$ and (b) the Compton wavelength $\lambda_C$ from the observed abundance of low-$z$ X-ray clusters, as functions of the index $n$. The medium-shaded band corresponds to the default limit reported in Schmidt et al. [30], whereas dark- and light-shaded areas use more or less conservative assumptions. Reproduced with permission from Ferraro et al. [34]. CL, confidence limit. (Online version in colour.)
significant future improvements possible. Clusters can also be used to test the relation between dynamical and lensing mass measures, which we will return to in §5.

The effects on halo clustering, quantified in terms of the linear halo bias, are generally consistent with what is expected from the abundance results: a higher halo abundance at fixed mass corresponds to a lower significance and hence a smaller bias at that mass. While the magnitude of the effect is smaller than that on the abundance ($\Delta b/b \sim 0.1–0.2$), future, larger cluster samples will also be able to make use of the clustering information to constrain modified gravity and break parameter degeneracies.

4. Matter power spectrum and weak lensing

The nonlinear matter power spectrum contains significant information on the growth history of structure. Unfortunately, this information is difficult to extract directly through the clustering of luminous tracers such as galaxies, due to the uncertainty in the biasing of these tracers. Weak lensing, on the other hand, circumvents these uncertainties and can serve as a probe of modifications to gravity [35–39].

Figure 4 shows the ratio of the nonlinear power spectrum measured in $f(R)$ simulations to that in $\Lambda$CDM simulations. The shaded region shows the predictions of a halo model approach [33], using the two cases of spherical collapse. In the halo model, the matter power spectrum is a sum of one- and two-halo terms. The latter is essentially given by the clustering of halos

---

*Phil. Trans. R. Soc. A* (2011)
weighted by bias and dominates on large scales, where it asymptotes to the linear matter power spectrum. The one-halo term is made up of the density profiles of halos and dominates on small scales. This results in a characteristic shape of the change in $P(k)$ predicted in the halo model, which peaks at around $k \sim 0.5h\text{Mpc}^{-1}$, where the one-halo term begins to dominate over the two-halo term. The simulations exhibit a similar behaviour for the large and intermediate field values, though the amplitude of the $P(k)$ enhancement is somewhat smaller than predicted.

The impact of the chameleon mechanism is seen in the intermediate and especially the small-field cases. Since our spherical collapse calculation does not include that effect, it is not surprising that the small-field case is not matched by the halo model prediction. Here, a more careful spherical collapse study solving for the $f_R$ field consistently will be necessary to make progress. Qualitatively, the impact of the chameleon mechanism is as expected: since massive halos are the regions most likely to be screened, we expect their contribution to the one-halo term to be affected first, which impacts the power spectrum on scales $0.4 \lesssim k \lesssim 1h\text{Mpc}^{-1}$.

Figure 5 shows the relative suppression of the power spectrum in the sDGP model. Comparing the linearized DGP simulations to the full simulations, we clearly see the effect of the Vainshtein mechanism. Remarkably, our spherical collapse calculation coupled with the halo model describes the simulation results very well: the linearized DGP spherical collapse result matches the linearized DGP simulations (up to $k \sim 1.5h\text{Mpc}^{-1}$), while the predictions based on the
Figure 6. Relative deviation of the shear power spectrum in an $f(R)$ model with $|f_{R0}| = 10^{-6}$ from $\Lambda$CDM as function of multi-pole $\ell$. The triangles (Hu-Sawicki model, $f_{R0} = 10^{-6}$, linear) show the linear theory prediction, whereas the squares ($N$-body result, same) use the power spectrum measured as a function of redshift in the full $f(R)$ simulations. Solid line: $\Lambda$CDM, nonlinear $P(k)$.

Reproduced with permission from Schmidt [36]. (Online version in colour.)

tophat collapse including the Vainshtein mechanism match the full DGP results very well. Again, we find that the DGP and the Vainshtein mechanisms are easier to model analytically.

With upcoming surveys such as Dark Energy Survey, PanStarrs, Subaru Measurement of Images and Redshifts, Large Synoptic Survey Telescope (LSST) and EUCLID, weak lensing measurements of the nonlinear matter power spectrum will be a powerful probe of modified gravity. This is exemplified by figure 6, which shows projected error bars on the shear auto-correlation $C_{kk}(\ell)$ relative to $\Lambda$CDM expected for LSST [36]. For this, we assumed a sky fraction covered of 0.5, a total source galaxy density of 50 arcmin$^{-2}$, and considered source galaxies in a redshift slice of width $\Delta z \approx 0.8$ centred at $\bar{z} = 2.4$. We show the error bars superimposed on the prediction of $C_{kk}(\ell)$ for the small-field $f(R)$ model in linear theory (triangles), and using the full $N$-body measurement of $P(k, z)$ (squares). Clearly, it is crucial to take into account nonlinear $f(R)$ effects; however, LSST should be able to easily constrain $f(R)$ gravity in a regime where cluster abundance is not an effective probe anymore (§3). Similar expectations hold for the first generation of 21 cm surveys [40].

5. Dynamics

In a scalar–tensor theory, the physical metric can be written as $g_{\mu\nu} = \exp(2\phi)\tilde{g}_{\mu\nu}$, where $\phi$ is the scalar degree of freedom and $\tilde{g}_{\mu\nu}$ is the Einstein frame metric, sourced in the usual way by the matter stress-energy. For a zero rest mass
particle such as the photon, the geodesic is thus not affected directly by the scalar field [41]. In other words, the ‘lensing potential’ $\Phi_\pm = (\Phi - \Psi)/2$ is unchanged in the presence of $\phi$ and masses inferred from lensing are the same as those that would be inferred in GR given the same mass distribution.\footnote{Apart from changes in the gravitational constant, which can, in general, vary in time and space. For the models considered here, this effect is negligible ($f(R)$) or absent (DGP).} On the other hand, the gravitational force experienced by non-relativistic particles is given by the gradient of the potential $\Phi$, which receives a contribution from the scalar field (equation (2.5)). This directly affects masses $M_{\text{dyn}}$ inferred from dynamical measures, such as velocities and virial temperatures. Thus, comparing dynamical mass estimates of objects such as galaxies and clusters with their lensing-inferred mass can be used as a generic test for modifications to gravity [42–44].

As an example, we plot the force–weighted force modification $g_{\text{vir}} [44]$, measured within dark matter halos up to the virial radius, as a function of halo mass in $f(R)$ gravity in figure 7. As shown in the study of Schmidt [44], commonly used dynamical mass measures defined in terms of a threshold overdensity criterion will be modified as $M_{\text{dyn}} = g^{3/5} M$, where $M$ is the lensing mass. The points in figure 7 show measurements for individual halos in the simulations, while the solid lines are what is expected from a spherically symmetric solution of the $f_R$ field equation assuming a Navarro–Frenk–White density profile. We see that in the large-field $f(R)$ model, essentially all halos are unscreened and $g_{\text{vir}} = 4/3$, as expected from the discussion in §2. For the small-field case, all halos below $\sim 10^{14} M_\odot/h$ are screened. For the intermediate field value ($|f_{R0}| = 10^{-5}$), there is a transition from unscreened to screened in the range $10^{14} - 10^{15} M_\odot/h$. These trends

Figure 7. $g_{\text{vir}}$ measured for well-resolved halos in $f(R)$ simulations (points). The simulation results match the spherically symmetric calculations (lines). The circled points are halos that have a more massive halo in their immediate vicinity (see text). Reproduced with permission from Schmidt [44]. (Online version in colour.)
are very well matched by the spherical calculation. Interestingly, there are several partially screened low-mass halos ($<10^{14} M_\odot/h$) in the intermediate field case that according to the spherically symmetric solution should not be screened. Almost all of these outliers are located near more massive halos (circled points), showing the environmental dependence of the screening mechanism (‘blanket screening’).

Thus, care must be taken when using actual measurements of dynamical masses to constrain modified gravity of the chameleon type. Ideally, one would like to choose isolated objects for this test, in order to exclude blanket screening effects.

Again, the Vainshtein mechanism leads to a different behaviour in the DGP model: as it operates above a fixed value of enclosed matter density, $g_{\text{vir}}$ is not mass dependent, but depends on the details of the halo profile [44].

6. Conclusion

With the advent of precision cosmology, large-scale structure has become an increasingly sensitive tool for testing gravity. Viable modified gravity models however need to employ a mechanism to restore GR locally in order to pass solar system tests. Such a mechanism will generally affect large-scale structure observables on scales where the density field becomes nonlinear. We have presented results on the abundance of massive halos, the nonlinear matter power spectrum and the dynamics within halos for two particular models, $f(R)$ gravity and the DGP model, which represent different classes of nonlinear screening mechanisms. While the screening depends strongly on mass and environment in chameleon theories such as $f(R)$ gravity, it operates at a fixed density threshold in the case of the Vainshtein mechanism for the DGP model. This leaves characteristic imprints in the abundance and dynamics of clusters as well as the matter power spectrum. The abundance of clusters has already been used to rule out a sizeable portion of the $f(R)$ parameter space, with future improvements by an order of magnitude possible. To constrain $f(R)$ models further, it will be necessary to consider smaller objects (groups), as massive clusters will be screened. We also anticipate that the tightest future constraints on the DGP model and its generalizations will come from large-scale structure in the nonlinear regime.

This work was supported by the Gordon and Betty Moore Foundation at Caltech. The simulations used in this work have been performed on the Joint Fermilab—KICP Supercomputing Cluster, supported by grants from Fermilab, Kavli Institute for Cosmological Physics and the University of Chicago.

References

Review. Nonlinear modified gravity


32 Lombriser, L., Slosar, A., Seljak, U. & Hu, W. 2010 Constraints on f(R) gravity from probing the large-scale structure. (http://arxiv.org/abs/1003.3009v2)


42 Smith, T. L. 2009 Testing gravity on kiloparsec scales with strong gravitational lenses. (http://arxiv.org/abs/0907.4829v2)
